

Oy 1 ctccctgctgttgcgtcctccgcgagcagagat 42
Db 213 ctccctgctgttgcgtcctccgcgagcagagat 254

RESULT 10

US-07-772-979-1

Sequence 1, Application US/07772979

GENERAL INFORMATION:

APPLICANT: Mizoguchi, Junzo

APPLICANT: Nogawa, Makoto

APPLICANT: Yamashita, Tomoyuki

APPLICANT: Kubota, Akiko

APPLICANT: Maeda, Tadashi

APPLICANT: Taniguchi, Masahiko

APPLICANT: Kobayashi, Junji

APPLICANT: Satoh, Sakae

APPLICANT: Otsu, Masaru

TITLE OF INVENTION: STABLE PRODUCTION OF M-CSF ACTIVE

TITLE OF INVENTION: POLYPEPTIDE

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,

ADDRESS: P.C.

STREET: 1755 Jefferson Davis Highway, Fourth Floor

CITY: Arlington

STATE: Virginia

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/772.979

FILING DATE: 19911008

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Oblon, Norman F.

REGISTRATION NUMBER: 24,618

REFERENCE/DOCKET NUMBER: 2103-025-0

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703)521-4500

TELEFAX: (703)486-2347

TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 693 base pairs

TYPE: NUCLEIC ACID

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

IMMEDIATE SOURCE:

CLONE: M1-del

FEATURE:

NAME/KEY: CDS

LOCATION: 1..690

US-07-772-979-1

Query Match

Best Local Similarity 100.0%; Score 42; DB 3; Length 693;

Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ctccctgctgttgcgtcctccgcgagcagagat 42
ctccctgctgttgcgtcctccgcgagcagagat 92

US-08-426-571-3

Sequence 3, Application US/08426571

GENERAL INFORMATION:

APPLICANT: LADNER, MARTHA B.

APPLICANT: NOBLE, JANELLE A.

APPLICANT: MARTIN, GEORGE A.

APPLICANT: KAWASAKI, ERNEST S.

APPLICANT: COYNE, MAZIE YEE

APPLICANT: HALENBECK, ROBERT F.

APPLICANT: KOTHS, KIRSTON E.

TITLE OF INVENTION: NEW FORMS OF COLONY STIMULATING FACTOR-1

NUMBER OF SEQUENCES: 24

CORRESPONDENCE ADDRESS:

ADDRESSEE: CHIRON CORPORATION

STREET: Intellectual Property - R440, P.O. Box 8097

CITY: Emeryville

STATE: California

COUNTRY: U.S.A.

ZIP: 94662-8097

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/426.571

FILING DATE: 21-SEP-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Mcgarrittle Jr., Phillip L.

REGISTRATION NUMBER: 31,395

REFERENCE/DOCKET NUMBER: 0681.015

TELECOMMUNICATION INFORMATION:

TELEPHONE: (510) 601-2718

TELEFAX: (510) 655-3542

TELEX: n/a

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 1642 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS

LOCATION: 179..946

FEATURE:

NAME/KEY: mat_peptide

LOCATION: 275..946

FEATURE:

NAME/KEY: -

LOCATION: 340

OTHER INFORMATION: /note= "Intron Sequence"

US-08-426-571-3

Query Match

Best Local Similarity 100.0%; Score 42; DB 8; Length 1642;

Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ctccctgctgttgcgtcctccgcgagcagagat 42
Db 229 ctccctgctgttgcgtcctccgcgagcagagat 270

RESULT 12

US-60-278-1861

Sequence 1861, Application US/60278258

GENERAL INFORMATION:

APPLICANT: Morris, Macdonald

APPLICANT: Lal, Preeti

APPLICANT: Diep, Dinh

TITLE OF INVENTION: Method for the Identification of Sequence Polymorphism Using

OY 1 ctccctgtctgtgtgtgtctctctgcgcagcagagat 42
|||||
Db 213 ctccctgtctgtgtgtgtctctctgcgcagcagagat 254

RESULT 10
US-07-772-979-1

Sequence 1, Application US/07772979
GENERAL INFORMATION:
APPLICANT: Mizoguchi, Junzo
APPLICANT: Nogawa, Makoto
APPLICANT: Yamashita, Tomoyuki
APPLICANT: Kubota, Akiko
APPLICANT: Maeda, Tadashi
APPLICANT: Tanuchi, Masahiko
APPLICANT: Kobayashi, Junji
APPLICANT: Otsu, Masaru
TITLE OF INVENTION: STABLE PRODUCTION OF M-CSF ACTIVE
TITLE OF INVENTION: POLYPEPTIDE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESS: P.C.
STREET: 1755 Jefferson Davis Highway, Fourth Floor
CITY: Arlington
STATE: Virginia
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/772,979
FILING DATE: 1991.10.08
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Oblon, Norman F.
REGISTRATION/DOCKET NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 2103-025-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)521-4500
TELEFAX: (703)486-2347
TELEX: 248855 OPT UR
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 693 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: M1-del
FEATURE:
NAME/KEY: CDS
LOCATION: 1..690
US-07-772-979-1

Query Match 100.0%; Score 42; DB 3; Length 693;
Best Local Similarity 100.0%; Pred. No. 0.0001;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ctccctgtctgtgtgtgtctctctgcgcagcagagat 42
|||||
ctccctgtctgtgtgtgtctctctgcgcagcagagat 92

US-08-426-571-3

Sequence 3, Application US/08426571
GENERAL INFORMATION:
APPLICANT: LADNER, MARTHA B.
APPLICANT: NOBLE, JANELLE A.
APPLICANT: MARTIN, GEORGE A.
APPLICANT: KAWASAKI, ERNEST S.
APPLICANT: COYNE, MAZIE YEE
APPLICANT: HALENECK, ROBERT F.
APPLICANT: KOTHS, KIRSTON E.
TITLE OF INVENTION: NEW FORMS OF COLONY STIMULATING FACTOR-1
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: CHIRON CORPORATION
STREET: Intellectual Property - R440, P.O. Box 8097
CITY: Emeryville
STATE: California
COUNTRY: U.S.A.
ZIP: 94662-8097
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/426,571
FILING DATE: 21-APR-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mcgarigle Jr., Phillip L.
REGISTRATION/DOCKET NUMBER: 31,395
REFERENCE/DOCKET NUMBER: 0681.015
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2718
TELEFAX: (510) 655-3542
TELEX: n/a
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1642 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 179..946
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 275..946
FEATURE:
NAME/KEY: -
LOCATION: 340
OTHER INFORMATION: /note= "Intron Sequence"
US-08-426-571-3

Query Match 100.0%; Score 42; DB 8; Length 1642;
Best Local Similarity 100.0%; Pred. No. 0.0001;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctccctgtctgtgtgtgtctctctgcgcagcagagat 42
|||||
Db 229 ctccctgtctgtgtgtgtctctctgcgcagcagagat 270

RESULT 12
US-60-278-1861

Sequence 1861, Application US/60278258
GENERAL INFORMATION:
APPLICANT: Morris, MacDonald
APPLICANT: Lal, Preeti
APPLICANT: Diep, Dinh
TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using

→ US PN 6156300

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 11, 2002, 03:52:43 : Search time 2666.35 Seconds
(without alignments)
340.790 Million cell updates/sec

Title: US-09-786-214-11
Perfect score: 42
Sequence: 1 ctccctgctgtgtgtgtc.....ctccgtgcagcagagatc 42

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 21979536 seqs, 10817449327 residues

Total number of hits satisfying chosen parameters: 43959072

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Pending_Patents_NA_Main:*

1: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
8: /cgn2_6/ptodata/2/pna/US092_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US093_COMB.seq:*
10: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US095_COMB.seq:*
12: /cgn2_6/ptodata/2/pna/US096_COMB.seq:*
13: /cgn2_6/ptodata/2/pna/US097_COMB.seq:*
14: /cgn2_6/ptodata/2/pna/US098_COMB.seq:*
15: /cgn2_6/ptodata/2/pna/US099_COMB.seq:*
16: /cgn2_6/ptodata/2/pna/US100_COMB.seq:*
17: /cgn2_6/ptodata/2/pna/US101_COMB.seq:*
18: /cgn2_6/ptodata/2/pna/US102_COMB.seq:*
19: /cgn2_6/ptodata/2/pna/US103_COMB.seq:*
20: /cgn2_6/ptodata/2/pna/US104_COMB.seq:*
21: /cgn2_6/ptodata/2/pna/US105_COMB.seq:*
22: /cgn2_6/ptodata/2/pna/US106_COMB.seq:*
23: /cgn2_6/ptodata/2/pna/US107_COMB.seq:*
24: /cgn2_6/ptodata/2/pna/US108_COMB.seq:*
25: /cgn2_6/ptodata/2/pna/US109_COMB.seq:*
26: /cgn2_6/ptodata/2/pna/US110_COMB.seq:*
27: /cgn2_6/ptodata/2/pna/US111_COMB.seq:*
28: /cgn2_6/ptodata/2/pna/US112_COMB.seq:*
29: /cgn2_6/ptodata/2/pna/US113_COMB.seq:*
30: /cgn2_6/ptodata/2/pna/US114_COMB.seq:*
31: /cgn2_6/ptodata/2/pna/US115_COMB.seq:*
32: /cgn2_6/ptodata/2/pna/US116_COMB.seq:*
33: /cgn2_6/ptodata/2/pna/US117_COMB.seq:*
34: /cgn2_6/ptodata/2/pna/US118_COMB.seq:*
35: /cgn2_6/ptodata/2/pna/US119_COMB.seq:*
36: /cgn2_6/ptodata/2/pna/US120_COMB.seq:*
37: /cgn2_6/ptodata/2/pna/US121_COMB.seq:*
38: /cgn2_6/ptodata/2/pna/US122_COMB.seq:*
39: /cgn2_6/ptodata/2/pna/US123_COMB.seq:*
40: /cgn2_6/ptodata/2/pna/US124_COMB.seq:*
41: /cgn2_6/ptodata/2/pna/US125_COMB.seq:*
42: /cgn2_6/ptodata/2/pna/US126_COMB.seq:*
43: /cgn2_6/ptodata/2/pna/US127_COMB.seq:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	100.0	42	1 PCT-US99-20344-11	Sequence 11, Appl
2	42	100.0	42	US-09-786-214-11	Sequence 11, Appl
3	42	100.0	60	1 PCT-US99-20344-8	Sequence 8, Appl
4	42	100.0	60	US-09-786-214-8	Sequence 8, Appl
5	42	100.0	78	1 PCT-US99-20344-4	Sequence 4, Appl
6	42	100.0	78	US-09-786-214-4	Sequence 4, Appl
7	42	100.0	129	1 PCT-US99-20344-50	Sequence 50, Appl
8	42	100.0	129	US-09-786-214-50	Sequence 50, Appl
9	42	100.0	609	29 US-09-758-449-148	Sequence 148, Appl
10	42	100.0	693	3 US-07-772-979-11	Sequence 1, Appl
11	42	100.0	1642	8 US-08-426-571-3	Sequence 3, Appl
12	42	100.0	1650	6 US-60-278-258-1861	Sequence 1861, Appl
13	42	100.0	1994	37 US-10-044-080-566	Sequence 566, Appl
14	42	100.0	1994	65 US-60-260-483-566	Sequence 566, Appl
15	42	100.0	1994	66 US-60-278-258-1859	Sequence 1859, Appl
16	42	100.0	2237	3 US-07-939-295-1	Sequence 1, Appl
17	42	100.0	2482	56 US-60-172-373-15186	Sequence 15186, Appl
18	42	100.0	2771	66 US-60-278-258-1863	Sequence 1863, Appl
19	42	100.0	3391	58 US-60-195-106-52	Sequence 52, Appl
20	42	100.0	3640	56 US-60-209-106-415	Sequence 415, Appl
21	42	100.0	3640	55 US-60-172-373-1570	Sequence 1570, Appl
22	42	100.0	3939	18 US-09-442-589B-619	Sequence 619, Appl
23	42	100.0	6306	4 US-08-004-141A-1	Sequence 1, Appl
24	40	95.2	2302	8 US-08-426-571-1	Sequence 1, Appl
25	32.4	77.1	320	28 US-09-704-424-28321	Sequence 28321, Appl
26	32.4	77.1	338	28 US-09-704-424-28722	Sequence 28722, Appl
27	32.4	77.1	369	28 US-09-704-424-24902	Sequence 24902, Appl
28	32.4	77.1	1987	8 US-08-426-571-23	Sequence 23, Appl
29	32.4	77.1	3920	75 US-60-360-207-11684	Sequence 11684, Appl
30	32.4	77.1	3931	8 US-08-426-571-21	Sequence 21, Appl
31	30.8	73.3	319	28 US-09-704-424-30476	Sequence 30476, Appl

```
32 25 59.5 364 31 US-09-821-837-5558 Sequence 5558, Ap
33 25 59.5 479 25 US-09-649-163-7958 Sequence 7958, Ap
34 25 59.5 479 25 US-09-652-911-7860 Sequence 7860, Ap
35 25 59.5 488 18 US-09-474-436-629 Sequence 629, Ap
36 25 59.5 488 25 US-09-649-163-5318 Sequence 5318, Ap
37 25 59.5 489 18 US-09-474-436-237 Sequence 237, Ap
38 25 59.5 489 25 US-09-649-163-5283 Sequence 5283, Ap
39 25 59.5 489 25 US-09-652-911-1444 Sequence 1444, Ap
40 25 59.5 495 18 US-09-474-436-6152 Sequence 6152, Ap
41 25 59.5 644 25 US-09-649-163-10225 Sequence 10225, A
42 25 59.5 206025 63 US-60-243-468-603 Sequence 603, App
43 24.6 58.6 451 33 US-09-873-402A-36830 Sequence 36830, A
44 24.6 58.6 451 59 US-60-209-830-36692 Sequence 36692, A
c 45 23.4 55.7 5061 24 US-09-620-392-23282 Sequence 23282, A
```

ALIGNMENTS

```
RESULT 1
PCT-US99-20344-11
; Sequence 11, Application PC/TUS9920344
; GENERAL INFORMATION:
; APPLICANT: Ludwig Institute for Cancer Research
; TITLE OF INVENTION: AN ANTIGENIC PEPTIDE ENCODED BY AN ALTERNATIVE OPEN READING FRAME
; FILE REFERENCE: L0461/7040MO
; CURRENT APPLICATION NUMBER: PCT/US99/20344
; CURRENT FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: FastSeq for Window Version 3.0
; SEQ ID NO 11
; LENGTH: 42
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..42
PCT-US99-20344-11
```

```
Query Match 100.0%; Score 42; DB 1; Length 42;
Best Local Similarity 100.0%; Pred. No. 7.5e-05;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Db 1 ctccctgtgtgtgtgtgtctctccctgagcagagagat 42
|||
1 ctccctgtgtgtgtgtgtctctccctgagcagagagat 42

RESULT 2
US-09-786-214-11
; Sequence 11, Application US/09786214
; GENERAL INFORMATION:
; APPLICANT: Ludwig Institute for Cancer Research
; TITLE OF INVENTION: AN ANTIGENIC PEPTIDE ENCODED BY AN ALTERNATIVE OPEN READING FRAME
; FILE REFERENCE: L0461/7040MO
; CURRENT APPLICATION NUMBER: US/09/786,214
; CURRENT FILING DATE: 2001-06-14
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: FastSeq for Window Version 3.0
; SEQ ID NO 11
; LENGTH: 42
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..42
US-09-786-214-11
```

Query Match 100.0%; Score 42; DB 30; Length 42;

Best Local Similarity 100.0%; Pred. No. 7.5e-05; Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 ctccctgtgtgtgtgtgtctctccctgagcagagagat 42
|||
Db 1 ctccctgtgtgtgtgtgtctctccctgagcagagagat 42
```

```
RESULT 3
PCT-US99-20344-8
; Sequence 8, Application PC/TUS9920344
; GENERAL INFORMATION:
; APPLICANT: Ludwig Institute for Cancer Research
; TITLE OF INVENTION: AN ANTIGENIC PEPTIDE ENCODED BY AN ALTERNATIVE OPEN READING FR
; FILE REFERENCE: L0461/7040MO
; CURRENT APPLICATION NUMBER: PCT/US99/20344
; CURRENT FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: FastSeq for Window Version 3.0
; SEQ ID NO 8
; LENGTH: 60
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..60
PCT-US99-20344-8
```

Query Match 100.0%; Score 42; DB 1; Length 60; Best Local Similarity 100.0%; Pred. No. 7.8e-05; Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 ctccctgtgtgtgtgtgtctctccctgagcagagagat 42
|||
Db 10 ctccctgtgtgtgtgtgtctctccctgagcagagagat 51
```

```
RESULT 4
US-09-786-214-8
; Sequence 8, Application US/09786214
; GENERAL INFORMATION:
; APPLICANT: Ludwig Institute for Cancer Research
; TITLE OF INVENTION: AN ANTIGENIC PEPTIDE ENCODED BY AN ALTERNATIVE OPEN READING FR
; FILE REFERENCE: L0461/7040MO
; CURRENT APPLICATION NUMBER: US/09/786,214
; CURRENT FILING DATE: 2001-06-14
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: FastSeq for Window Version 3.0
; SEQ ID NO 8
; LENGTH: 60
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..60
US-09-786-214-8
```

Query Match 100.0%; Score 42; DB 30; Length 60; Best Local Similarity 100.0%; Pred. No. 7.8e-05; Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 ctccctgtgtgtgtgtgtctctccctgagcagagagat 42
|||
Db 10 ctccctgtgtgtgtgtgtctctccctgagcagagagat 51
```

RESULT 5
PCT-US99-20344-4
; Sequence 4, Application PC/TUS9920344

```
; GENERAL INFORMATION:
; APPLICANT: Ludwig Institute for Cancer Research
; TITLE OF INVENTION: AN ANTIGENIC PEPTIDE ENCODED BY AN ALTERNATIVE OPEN READING FRAME
; FILE REFERENCE: L0461/7040MO
; CURRENT APPLICATION NUMBER: PCT/US99/20344
; CURRENT FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: FastSeq for Window Version 3.0
; SEQ ID NO 4
; LENGTH: 78
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..75
PCT-US99-20344-4

Query Match          100.0%; Score 42; DB 1; Length 78;
Best Local Similarity 100.0%; Pred. No. 8e-05;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctccctgctgtgtgtgtctgtctctcctcgagcagagat 42
    |||||||
Db 10 ctccctgctgtgtgtgtctgtctctcctcgagcagagat 51

RESULT 6
US-09-786-214-4
; Sequence 4, Application US/09786214
; GENERAL INFORMATION:
; APPLICANT: Ludwig Institute for Cancer Research
; TITLE OF INVENTION: AN ANTIGENIC PEPTIDE ENCODED BY AN ALTERNATIVE OPEN READING FRAME
; FILE REFERENCE: L0461/7040MO
; CURRENT APPLICATION NUMBER: US/09/786,214
; CURRENT FILING DATE: 2001-06-14
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: FastSeq for Window Version 3.0
; SEQ ID NO 4
; LENGTH: 78
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..75
US-09-786-214-4

Query Match          100.0%; Score 42; DB 30; Length 78;
Best Local Similarity 100.0%; Pred. No. 8e-05;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctccctgctgtgtgtgtctgtctctcctcgagcagagat 42
    |||||||
Db 10 ctccctgctgtgtgtgtctgtctctcctcgagcagagat 51

RESULT 7
PCT-US99-20344-50
; Sequence 50, Application PC/TUS9920344
; GENERAL INFORMATION:
; APPLICANT: Ludwig Institute for Cancer Research
; TITLE OF INVENTION: AN ANTIGENIC PEPTIDE ENCODED BY AN ALTERNATIVE OPEN READING FRAME
; FILE REFERENCE: L0461/7040MO
; CURRENT APPLICATION NUMBER: PCT/US99/20344
; CURRENT FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: FastSeq for Window Version 3.0
; SEQ ID NO 50
; LENGTH: 129
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..129
PCT-US99-20344-50

Query Match          100.0%; Score 42; DB 1; Length 129;
Best Local Similarity 100.0%; Pred. No. 8.5e-05;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctccctgctgtgtgtgtctgtctctcctcgagcagagat 42
    |||||||
Db 51 ctccctgctgtgtgtgtctgtctctcctcgagcagagat 92

RESULT 8
US-09-786-214-50
; Sequence 50, Application US/09786214
; GENERAL INFORMATION:
; APPLICANT: Ludwig Institute for Cancer Research
; TITLE OF INVENTION: AN ANTIGENIC PEPTIDE ENCODED BY AN ALTERNATIVE OPEN READING FR
; FILE REFERENCE: L0461/7040MO
; CURRENT APPLICATION NUMBER: US/09/786,214
; CURRENT FILING DATE: 2001-06-14
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: FastSeq for Window Version 3.0
; SEQ ID NO 50
; LENGTH: 129
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..129
US-09-786-214-50

Query Match          100.0%; Score 42; DB 30; Length 129;
Best Local Similarity 100.0%; Pred. No. 9.9e-05;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctccctgctgtgtgtgtctgtctctcctcgagcagagat 42
    |||||||
Db 51 ctccctgctgtgtgtgtctgtctctcctcgagcagagat 92

RESULT 9
US-09-758-449-148
; Sequence 148, Application US/09758449
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PM026
; CURRENT APPLICATION NUMBER: US/09/758,449
; CURRENT FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 1478
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 148
; LENGTH: 609
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-758-449-148

Query Match          100.0%; Score 42; DB 29; Length 609;
Best Local Similarity 100.0%; Pred. No. 9.9e-05;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```



```
; TITLE OF INVENTION: Polynucleotide Sequence Databases, and Single Nucleotide
; FILE OF INVENTION: Polymorphisms Identified Thereby
; FILE REFERENCE: GX-0010-1 P
; CURRENT APPLICATION NUMBER: US/60/278,258
; NUMBER OF SEQ ID NOS: 17730
; SOFTWARE: PERL Program
; SEQ ID NO 1861
; LENGTH: 1690
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 333948.12
US-60-278-258-1861

Query Match          100.0%; Score 42; DB 66; Length 1690;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctccctgctgtgtgtgtctgtctctcctcgcgcagcagagat 42
    |||||||
Db 470 ctccctgctgtgtgtgtctgtctctcctcgcgcagcagagat 511

RESULT 13
US-10-044-090-566
; Sequence 566, Application US/10044090
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 566
; LENGTH: 1994
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 1985219CB1
US-10-044-090-566

Query Match          100.0%; Score 42; DB 37; Length 1994;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctccctgctgtgtgtgtctgtctctcctcgcgcagcagagat 42
    |||||||
Db 207 ctccctgctgtgtgtgtctgtctctcctcgcgcagcagagat 248

RESULT 14
US-60-260-483-566
; Sequence 566, Application US/60260483
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 P
; CURRENT APPLICATION NUMBER: US/60/260,483
; CURRENT FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 566
; LENGTH: 1994
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
```

```
; OTHER INFORMATION: Incyte ID No: 1985219CB1
US-60-260-483-566

Query Match          100.0%; Score 42; DB 65; Length 1994;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctccctgctgtgtgtgtctgtctctcctcgcgcagcagagat 42
    |||||||
Db 207 ctccctgctgtgtgtgtctgtctctcctcgcgcagcagagat 248

RESULT 15
US-60-278-258-1859
; Sequence 1859, Application US/60278258
; GENERAL INFORMATION:
; APPLICANT: Morris, MacDonald
; APPLICANT: Lal, Preeti
; APPLICANT: Diep, Dinh
; TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using
; FILE OF INVENTION: Polynucleotide Sequence Databases, and Single Nucleotide
; FILE REFERENCE: GX-0010-1 P
; CURRENT APPLICATION NUMBER: US/60/278,258
; CURRENT FILING DATE: 2001-03-23
; NUMBER OF SEQ ID NOS: 17730
; SOFTWARE: PERL Program
; SEQ ID NO 1859
; LENGTH: 1994
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 333948.11
US-60-278-258-1859

Query Match          100.0%; Score 42; DB 66; Length 1994;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctccctgctgtgtgtgtctgtctctcctcgcgcagcagagat 42
    |||||||
Db 207 ctccctgctgtgtgtgtctgtctctcctcgcgcagcagagat 248
```

Search completed: September 11, 2002, 05:37:54
Job time: 6311 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2002, 06:46:47 : Search time 57.95 Seconds
(without alignments)
23.214 Million cell updates/sec

Title: US-09-786-214-12
Perfect score: 72
Sequence: 1 LPAVGLSPGREQ 14

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	61.1	381	2	cytochrome P450 -
2	43	59.7	211	2	precortin-6Y methy
3	43	59.7	623	2	probable lysophosp
4	43	59.7	821	2	DNA helicase (limp
5	41	56.9	156	2	conserved hypotet
6	41	56.9	326	2	probable N5, N10-me
7	41	56.9	508	2	probable cobi prot
8	40	55.6	227	2	hypothetical prote
9	40	55.6	243	2	MHC H-2K1-k - mous
10	40	55.6	428	2	uracil permease ho
11	40	55.6	428	2	uracil permease ho
12	40	55.6	540	2	carboxylesterase,
13	40	55.6	661	2	hypothetical prote
14	40	55.6	673	2	probable lysophosp
15	39	54.2	43	2	Ig kappa chain V r
16	39	54.2	96	2	Ig kappa chain V r
17	39	54.2	103	2	Ig kappa chain V r
18	39	54.2	106	2	Ig kappa chain V r
19	39	54.2	106	2	Ig kappa chain V r
20	39	54.2	107	2	Ig kappa chain (an
21	39	54.2	108	2	Ig kappa chain V-J
22	39	54.2	108	2	Ig kappa chain V r
23	39	54.2	108	2	Ig kappa chain V r
24	39	54.2	111	2	Ig kappa chain V r
25	39	54.2	114	2	Ig kappa chain V r
26	39	54.2	115	1	Ig kappa chain pre
27	39	54.2	115	1	Ig kappa chain pre
28	39	54.2	116	2	Ig kappa chain pre
29	39	54.2	116	2	Ig kappa chain pre

30	39	54.2	119	2	Ig kappa chain V r
31	39	54.2	125	2	Ig kappa chain V-J
32	39	54.2	128	2	Ig kappa chain pre
33	39	54.2	128	2	Ig kappa chain V-J
34	39	54.2	128	2	Ig kappa chain V r
35	39	54.2	129	2	Ig kappa chain V r
36	39	54.2	129	2	Ig kappa chain V r
37	39	54.2	132	2	Ig kappa chain pre
38	39	54.2	144	2	Ig kappa chain pre
39	39	54.2	144	2	Ig heavy chain V r
40	39	54.2	215	2	Ig kappa chain V-I
41	39	54.2	277	2	formyltetrahydrofo
42	39	54.2	278	2	formyltetrahydrofo
43	39	54.2	319	2	probable regulator
44	39	54.2	563	2	nerve growth facto
45	39	54.2	1367	2	BUD3 protein - yea

ALIGNMENTS

RESULT 1
F75270
Cytochrome P450 - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: F75270

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MID:20036896
A:Accession: F75270

A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-381 <MW>

A:Cross-references: GB:AE002076; GB:AE000513; MID:96460285; PIDN:AAF12016.1; PID:9646
A:Experimental source: strain R1

C:Genetics:

A:Gene: DR2473

A:Map position: 1

Query Match 61.1% Score 44; DB 2; Length 381;
Best local Similarity 100.0% Pred. No. 12;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAVGLSP 9
DB 52 LPAVGLSP 60

RESULT 2

A64490
precortin-6Y methylase homolog - Methanococcus jannaschii

C:Species: Methanococcus jannaschii

C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000

R:Rault, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak

; Reich, C.I.; Overbeek, R.; Kirsness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodok,

ison, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.

Science 273, 1058-1073, 1996

A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Moese

A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc

A:Reference number: A64300; MID:96337999

A:Accession: A64490

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-211 <MW>

A:Cross-references: GB:U67593; GB:L77117; MID:92826427; PIDN:AAB9541.1; PID:91592152

C:Genetics:

A:Map position: FOR1500322-1500957

C:Superfamily: precorrin-6Y methylase

Query Match 59.7%; Score 43; DB 2; Length 211;
Best Local Similarity 54.5%; Pred. No. 9.3;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 4 VVGLSPGDEY 14
:||:|||||
Db 4 IVGIGPDREY 14

RESULT 3
T40991

probable lysophospholipase precursor - fission yeast (*Schizosaccharomyces pombe*)
C:Species: *Schizosaccharomyces pombe*
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C:Accession: T40991
R:Llyne, M.; Rajandream, M.A.; Barrell, B.G.; Volckaert, G.
submitted to the EMBL Data Library, March 1999
A:Reference number: 221962
A:Accession: T40991

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-623 <LYN>
A:Cross-references: EMBL:AL049559; PIDN:CAB40176.1; GSPDB:GN00068; SPDB:SPCC1450.09C
A:Experimental source: strain 972h-; cosmid c1450
C:Genetics:
A:Gene: SPDB:SPCC1450.09C
A:Map position: 3
C:Superfamily: yeast lysophospholipase

Query Match 59.7%; Score 43; DB 2; Length 623;
Best Local Similarity 69.2%; Pred. No. 29;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 2 PAVGLSPGDEY 14
||| ||| |||
Db 76 PASDGLSTGEORF 88

RESULT 4
C84304

DNA helicase [imported] - *Halobacterium* sp. NRC-1
C:Species: *Halobacterium* sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: C84304
R:Ng, W.V.; Kennedy, S.P.; Mahiriz, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leitbauer, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A:Title: Genome sequence of *Halobacterium* species NRC-1.
A:Reference number: A84160; MUID:20504483
A:Accession: C84304

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-821 <STO>
A:Cross-references: GB:AE004437; NID:q10580995; PIDN:AA619799.1; GSPDB:GN00138
C:Genetics:
A:Gene: hel

Query Match 59.7%; Score 43; DB 2; Length 821;
Best Local Similarity 90.0%; Pred. No. 38;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 AVVGLSPGDEQ 12
|||||||
Db 326 AVVGLSPGDEQ 335

RESULT 5
F87551

conserved hypothetical protein CC2439 [imported] - *Caulobacter crescentus*
C:Species: *Caulobacter crescentus*
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C:Accession: F87551
R:Nleman, W.C.; Feldblum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of *Caulobacter crescentus*.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: F87551
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-156 <STO>
A:Cross-references: GB:AE005673; NID:q13423984; PIDN:AAK24410.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC2439
C:Superfamily: Haemophilus influenzae conserved hypothetical protein HI0305

Query Match 56.9%; Score 41; DB 2; Length 156;
Best Local Similarity 80.0%; Pred. No. 15;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 3 AVVGLSPGDEQ 12
|||||
Db 18 AVVGLDPGERK 27

RESULT 6
T45226

probable N5,N10-methylene-tetrahydromethanopterin reductase (F420-dependent) [imported]
C:Species: *Methanobacterium tindarius*
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C:Accession: T45226
R:Westenberg, D.J.; Braune, A.; Ruppert, C.; Mueller, V.; Herzberg, C.; Gottschalk, G
submitted to the EMBL Data Library, September 1998
A:Description: The F420H2-dehydrogenase from *Methanobacterium tindarius*. Cloning of the f
A:Reference number: 222947
A:Accession: T45226

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-326 <WES>
A:Cross-references: EMBL:AF011519; PIDN:CAB56639.1
A:Experimental source: DSM 2278
C:Genetics:
A:Gene: ftdA

Query Match 56.9%; Score 41; DB 2; Length 326;
Best Local Similarity 70.0%; Pred. No. 32;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 3 AVVGLSPGDEQ 12
||:|||||
Db 88 AILGLSPGDEQ 97

RESULT 7
E70764

probable cobI protein - *Mycobacterium tuberculosis* (strain H37RV)
C:Species: *Mycobacterium tuberculosis*
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: E70764
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentles, S.; Hamlin, N.; Bolroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete geno
A:Reference number: A70500; MUID:98295987

A:Accession: E70764
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-508 <COL>
 A:Cross-references: GB:273966; GB:AL123456; NID:g3261577; PIDN:CAA98214.1; PID:e246996;
 A:Experimental source: strain H37Rv
 A:Genetics:
 A:gene: cobi

Query Match 56.9%; Score 41; DB 2; Length 508;
 Best Local Similarity 58.3%; Pred. No. 51;
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 3 AVGLSPGEQEV 14
 ||||| ||: ||:
 Db 250 AVVGLGPGDSW 261

RESULT 8
 B90400
 hypothetical protein cblE [imported] - Sulfolobus solfataricus
 C:Species: Sulfolobus solfataricus
 C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 15-Jun-2001
 C:Accession: B90400

R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, H. et al., R.A.; Ragan, M.A.; Senses, C.W.; Van der Oost, J.
 submitted to GenBank, April 2001

A:Description: Sulfolobus solfataricus complete genome.

A:Reference number: A99139

A:Accession: B90400

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-227 <KUR>

A:Cross-references: GB:AE006641; NID:g13815599; PIDN:AAK42457.1; GSPDB:GN00155

C:Genetics:

A:gene: cblE

C:Superfamily: precorrin-6Y methylase

Query Match 55.6%; Score 40; DB 2; Length 227;
 Best Local Similarity 54.5%; Pred. No. 33;
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 4 VVGLSPGEQEV 14
 ||||| ||: ||:
 Db 10 IVGVGPGDPEY 20

RESULT 9

MHC H-2K1-k - mouse

C:Species: Mus musculus (house mouse)

C:Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 05-Nov-1999

C:Accession: I54459

R:Watts, S.; Davis, A.C.; Goodenow, R.S.

Immunogenetics 29, 355-357, 1989

A:Title: Sequence analysis of the C3H H-2K1-k gene: Relationship to the H-2 loci.

A:Reference number: I54459; MUID:89233303

A:Accession: I54459

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-243 <RES>

A:Cross-references: GB:M27134; NID:g199435; PIDN:AAA39610.1; PID:g387456

C:Genetics:

A:introns: 22/1, 112/1

Query Match 55.6%; Score 40; DB 2; Length 243;
 Best Local Similarity 77.8%; Pred. No. 35;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 5 VGLSPGEQEV 13
 :|||||: |
 Db 216 LGLSPGEE 224

RESULT 10

AG1304

uracil permease homolog pyrP [imported] - Listeria monocytogenes (strain EGD-e)

C:Species: Listeria monocytogenes

C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001

C:Accession: AG1304

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec

; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsthl,

D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A:Authors: Kref, J.; Kuhn, M.; Kunst, F.; Kurapk, G.; Madueno, E.; Maltouram, A.;

Ok, C.; Schleuter, T.; Simoes, N.; Tlerez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla

A:Title: Comparative genomics of Listeria species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AG1304

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-428 <GLA>

A:Cross-references: GB:NC_003210; PIDN:CAC99917.1; PID:g16411293; GSPDB:GN00177

A:Experimental source: strain EGD-e

C:Genetics:

A:gene: pyrP

C:Superfamily: uracil transport protein uraA

Query Match 55.6%; Score 40; DB 2; Length 428;
 Best Local Similarity 70.0%; Pred. No. 63;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 LPAVVGSLPG 10
 :||: |||||
 Db 42 VPSVVGSLPG 51

RESULT 11

AG1676

uracil permease homolog pyrP [imported] - Listeria innocua (strain C11p11262)

C:Species: Listeria innocua

C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001

C:Accession: AG1676

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec

; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsthl,

D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A:Authors: Kref, J.; Kuhn, M.; Kunst, F.; Kurapk, G.; Madueno, E.; Maltouram, A.;

Ok, C.; Schleuter, T.; Simoes, N.; Tlerez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla

A:Title: Comparative genomics of Listeria species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AG1676

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-428 <GLA>

A:Cross-references: GB:AL592022; PIDN:CAC97183.1; PID:g16414454; GSPDB:GN00178

A:Experimental source: strain C11p11262

C:Genetics:

A:gene: pyrP

C:Superfamily: uracil transport protein uraA

Query Match 55.6%; Score 40; DB 2; Length 428;
 Best Local Similarity 70.0%; Pred. No. 63;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 LPAVVGSLPG 10
 :||: |||||
 Db 42 VPSVVGSLPG 51

RESULT 12

A:75250
 A:carboxylesterase, type B - Deinococcus radiodurans (strain R1)
 C:Species: Deinococcus radiodurans
 C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
 C:Accession: A75250
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A:Reference number: A75250; MUID:20036896
 A:Accession: A75250
 A:Molecule type: DNA
 A:Residues: 1-540 <EMBL>
 A:Cross-references: GB:AE002092; GB:AE000513; NID:96460455; PIDN:AAF12163.1; PID:9646045
 A:Experimental source: strain R1
 C:Genetics:
 A:Gene: DR2626
 A:Map position: 1
 C:Superfamily: cholinesterase; cholinesterase homology

Query Match 55.6%; Score 40; DB 2; Length 540;
 Best Local Similarity 70.0%; Pred. No. 80;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 Oy 2 PAVVGLSPGEQ 11
 | : : : : :
 Db 512 PAVGLARGE 521

RESULT 13

G84511
 hypothetical protein At2g13900 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: G84511
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vankar, S.E.; Unayam, L.; Tallon, L.; Eust, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487
 A:Accession: G84511
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-661 <STO>
 A:Cross-references: GB:AE002093; NID:96598598; PIDN:AAF18650.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g13900
 A:Map position: 2

Query Match 55.6%; Score 40; DB 2; Length 661;
 Best Local Similarity 53.8%; Pred. No. 99;
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 Oy 2 PAVVGLSPGEQ 14
 | : : : : :
 Db 437 PLTKIVFGEQ 449

RESULT 14

T50281
 probable lysophospholipase (EC 3.1.1.5) precursor SPAC977.09c [similarity] - fission yeast
 C:Species: Schizosaccharomyces pombe
 C>Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 19-Jan-2001
 C:Accession: T50281; T42738
 R:Zimmermann, W.; Wambutt, R.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.
 submitted to the EMBL Data Library, January 2000
 A:Reference number: Z25053
 A:Accession: T50281

A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-673 <ZIM>

A:Cross-references: EMBL:AL137130; NID:96742151; PIDN:CAR69631.1; PID:96742159; GSPDB
 A:Experimental source: strain 972h(-); cosmid c977
 R:Yoshitaka, S.; Kato, K.; Nakai, K.; Okayama, H.; Nojima, H.
 DNA Res. 4, 363-369, 1997
 A:Title: Identification of open reading frames in Schizosaccharomyces pombe cDNAs.
 A:Reference number: Z17323; MUID:98162722
 A:Accession: T42738
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 427-673 <YOS>
 A:Cross-references: EMBL:D89183; NID:91749573; PIDN:BAAL3845.1; PID:91749574
 A:Experimental source: strain PR745
 C:Genetics:
 A:Gene: SPDB:SPAC977.09c
 A:Map position: 1
 A:Introns: 651/3
 C:Function:
 A:Description: catalyzes the hydrolysis of 2-lysophosphatidylcholine to glycerophospho
 C:Superfamily: yeast lysophospholipase
 C:Keywords: carboxylic ester hydrolase

Query Match 55.6%; Score 40; DB 2; Length 673;
 Best Local Similarity 61.5%; Pred. No. 1e+02;
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 Oy 2 PAVVGLSPGEQ 14
 | : : : : :
 Db 83 PASEGLMEGQSY 95

RESULT 15

S21065
 Ig kappa chain V region (anti-RH(D)) - human (fragment)
 C:Species: Homo sapiens (man)
 C>Date: 22-Nov-1993 #sequence_revision 24-May-1996 #text_change 09-May-1997
 C:Accession: S21065
 R:Dionisi, A.; Lecroisey, A.; Henschen, A.; Rouger, P.; Kell, B.
 Protein Seq. Data Anal. 4, 317-318, 1991
 A:Title: Subgroup assignment of a human monoclonal anti-Rh(D) antibody.
 A:Reference number: S21065; MUID:92235344
 A:Accession: S21065
 A:Molecule type: protein
 A:Residues: 1-43 <DIO>
 C:Superfamily: immunoglobulin V region; immunoglobulin homology
 C:Keywords: heterotrimer; immunoglobulin

Query Match 54.2%; Score 39; DB 2; Length 43;
 Best Local Similarity 63.6%; Pred. No. 8.5;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 Oy 2 PAVVGLSPGEQ 12
 | : : : : :
 Db 8 PATLISPGER 18

Search completed: September 9, 2002, 06:51:26
 Job time: 279 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 11, 2002, 02:45:43 : Search time 1587.89 Seconds
(without alignments)
356.997 Million cell updates/sec

Title: US-09-786-214-11
Perfect score: 42
Sequence: 1 ctccctgctgtgtgtgtct.....ctccctgagagagagat 42

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
EST:*
1: em_estba:*
2: em_esthum:*
3: em_estluc:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_estl:*
10: gb_estl2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	42	100.0	250	10	BF848044
2	42	100.0	520	10	BF724253
3	42	100.0	540	9	AL556517
4	42	100.0	574	10	AL556517
5	32.4	77.1	564	10	BF374641
6	32.4	77.1	579	10	BE912798
7	32.4	77.1	665	10	BE912798
8	32.4	77.1	672	10	BE912798
9	32.4	77.1	691	10	BE912798
10	32.4	77.1	731	10	BE912798
11	32.4	77.1	777	10	BE912798
12	32.4	77.1	825	10	BE912798
13	25.2	60.0	423	10	BE912798
14	25.2	60.0	448	10	BE912798
15	25.2	60.0	615	10	BE912798
16	24	57.1	579	10	BE912798
17	23.8	56.7	653	10	BE912798

18	23.4	55.7	764	10	BE263821
19	23.2	55.2	1135	10	BE783578
20	23	54.8	413	12	B40890
21	23	54.8	661	10	B063201
22	23	54.8	934	10	CNS03CR
23	23	54.8	1569	12	AG070941
24	22.8	54.3	282	12	A2553140
25	22.8	54.3	412	12	A2702302
26	22.8	54.3	539	12	BE075725
27	22.8	54.3	735	9	A1261290
28	22.8	54.3	768	10	B1822715
29	22.6	53.8	1137	10	BE786185
30	22.4	53.3	230	12	A2720631
31	22.4	53.3	240	10	B1332004
32	22.4	53.3	327	10	BE497080
33	22.4	53.3	387	10	BE491273
34	22.4	53.3	406	10	B1127593
35	22.4	53.3	442	10	B1120788
36	22.4	53.3	611	9	AU244251
37	22.4	53.3	622	10	BE506570
38	22.4	53.3	627	10	BE490922
39	22.4	53.3	629	10	BE499014
40	22.4	53.3	637	10	BE495620
41	22.4	53.3	638	10	BE376989
42	22.4	53.3	683	10	BE488388
43	22.4	53.3	704	12	A2854829
44	22.4	53.3	706	10	BE283113
45	22.4	53.3	753	10	BE505238

ALIGNMENTS

RESULT 1
LOCUS BF848044
DEFINITION CM3-EN0048-031100-458-e08 EN0048 Homo sapiens CDNA, mRNA sequence;
ACCESSION BF848044
VERSION BF848044.1 GI:12235194
KEYWORDS EST.
SOURCE Human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 250)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
CONTACT: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=CM3&cl2=CM3-EN0048-031100-458-e08&cl3=2000-11-03&cl4=1)
Seq primer: puc 18 forward
High quality sequence start: 10
High quality sequence stop: 249.
Location/Qualifiers
1..250
/organism="Homo sapiens"

```

/db.xref="taxon:3606"
/clone.lib="EN0048"
/dep-stage="Adult"
/note="Organ: lung-normal; Vector: puc18; Site:1: SmaI; Site:2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 47 a 51 c 89 g 61 t 2 others
ORIGIN

```

Query Match	100.0%	Score	42	DB	10	Length	250
Best Local Similarity	100.0%	Pred. No.	0	00019			
Matches	42	Conservative	0	Mismatches	0	Indels	0
						Gaps	0

QY 1 ctccctgctgttctgtctgtctcctgagcagagatat 42
|||||
133 CTCCCTGCTGTTGTTGGTCTGTCTCCTGGCAGCAGGAGTAT 174
Db

RESULT	2
BF724253	
LOCUS	BF724253
DEFINITION	520 bp mRNA linear EST 05-JAN-2001
DESCRIPTION	bx02f10.y1 Human Iris cDNA (Un-normalized, unamplified): BX Homo sapiens cDNA clone bx02f10 5', mRNA sequence.

ACCESSION	BF724253
VERSION	BF724253.1
KEYWORDS	EST.
SOURCE	human.

ORGANISM	REFERENCE
Homo sapiens	1 (bases 1 to 520)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	Wistow,G.,J., Bernstein,S., Behal,A. and Smith,D.
Invest. Ophthalmol. Vis. Sci. 41 (2000) In press	NETBANK: EST analysis and bioinformatics for ocular
Contact: Wistow G	

Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078
Email: grameghe@helix.nih.gov
Plate: 02 row: f column: 10
Seq primer: M13RP reverse primer (AB1).

FEATURES
source

```
/organism="Homo sapiens"  
/db_xref="taxon:9606"  
/clone="bx02f10"  
/clone_lib="Human Iris cDNA (Un-normalized, unamplified)"
```

```

/define type="Irls"
/dev stage="Adult"
/lab host="EMDHL0B"
/note="Organ: Eye; Vector: pcMWSPORT6; Post-mortem iris
tissue was pooled from 10 individuals ranging in age from
4-80 years and RNA was extracted. From this pooled sample
an aliquot of 60µg of total RNA yielded 2.17µg of mRNA. A
directionally cloned cDNA library in the pcMWSPORT6 vector
was constructed at Life Technologies, essentially
following the protocols of the Superscript Plasmid System
full details of which are contained in the manufacturer's
instruction manual (http://www.lifetech.com/). First
strand synthesis was carried out using a Not I
primer-adaptor [5'-pGATTAATCTTAAATAGCGAGCGCGCC(7)15-3'
]. Not I/dlunt end inserts were cloned into the Not I/EcoRI
v sites in the vector. EST analysis was performed on the
unamplified library at the NIH Intramural Sequencing
center (NISC)"

```

BASE COUNT	103 a	157 c	163 g	97 t
ORIGIN				

Query Match	100.0%	Score 42	DB 10	length 520	
Best Local Similarity	100.0%	Pred. No.	0 00023		
Matches 42;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0	

Qy 1 ctccgcgcgttctgtgtcgtctgtctccgcgcgcagcagat 42
 |||||
 Db 207 ctccgcgcgttctgtgtcgtctgtctccgcgcgcagcagat 248

RESULT	3	540 bp	mRNA	linear	EST 16-FEB-2001
AL556517					
LOCUS					
DEFINITION					
	AL556517				
	AL556517	LTI.NFL006.Pl2	Homo sapiens	cdna clone	CS00K0061121 5
					prime, mRNA sequence.

ACCESSION	AL556517	
VERSION	AL556517.1	GI:12899263
KEYWORDS	EST.	
SOURCE	human.	

ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Homo sapiens					
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	1 (bases 1 to 540)	Jesse, J. and Polayes, D.	Full-length cDNA libraries and normalization	Unpublished (2001)	Contact: Genoscope

Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
SOURCE

1.540

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DK006YR21"
/clone_1lb="L1-NEU06_PL2"
/tissue_type="Placenta"
/notes="vector: pcwvSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-Oligo(dt) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pcwvSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

```

Query Match	100.0%	Score 42:	DB 9:	Length 540:
Best Local Similarity	100.0%	Pred. NC:	0.00023:	
Matches 42:	Conservative 0:	Indels 0:	Gaps 0:	

```
QY      1 ctccctgctgttttgctcgtctctgcagacgaagtat 42
          |||||
Db     159 cttccctgctgttttgctcgtctcctggcagacgaagtat 200
```

RESULT	4
LOCUS	BG424422
DEFINITION	BG424422 974 bp mRNA Linear EST 14-MAR-2001 602447739F1 NIH_MGC_14 Homo sapiens CDNA clone IMAGE:4586312 5', mRNA sequence.

ACCESSION	BG424422	
VERSION	BG424422.1	GI:13330928
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 974)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: DCTD/DTF
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLM1314 row: n column: 09
High quality sequence stop: 709.

FEATURES
source

1..974
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4586312"
/clone_lib="NIH-MGC_14"
/tissue_type="renal cell adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: kidney; Vector: pOTB7; site_1: XhoI; site_2:
EcoRI; CDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAAGGAG(C). Size-selected >800bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-CDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT
199 a 279 c 299 g 196 t 1 others

ORIGIN

Query Match 100.0%; Score 42; DB 10; Length 974;
Best Local Similarity 100.0%; Pred. No. 0.00027;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctccctgctgtgtgtgtctctctcgcgcagcagagat 42
|||||
Db 241 ctccctgctgtgtgtgtctctcgcgcagcagagat 282
|||||

RESULT 5
BE374641 564 bp mRNA linear EST 21-JUL-2000
LOCUS 601224139F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3582536 5',
DEFINITION mRNA sequence.
ACCESSION BE374641
VERSION BE374641.1 GI:9320004
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 564)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLM8737 row: n column: 09
High quality sequence stop: 551.

FEATURES
source

Location/Qualifiers
1..564
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3582536"
/clone_lib="NCI_CGAP_Mam1"
/tissue_type="tumor, biopsy sample"
/dev_stage="10 months, virgin"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; site_1: SalI;
site_2: NotI; Cloned unidirectionally. Primer: Oligo dT
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

BASE COUNT
152 a 153 c 138 g 121 t

ORIGIN

Query Match 77.1%; Score 32.4; DB 10; Length 564;
Best Local Similarity 85.7%; Pred. No. 0.43;
Matches 36; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 ctccctgctgtgtgtgtctctcgcgcagcagagat 42
|||||
Db 40 ctccctgctgtgtgtgtctctcgcgcagcagagat 81
|||||

RESULT 6
BE912798 579 bp mRNA linear EST 29-SEP-2000
LOCUS 601664754F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3964933 5',
DEFINITION mRNA sequence.
ACCESSION BE912798
VERSION BE912798.1 GI:10409785
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 579)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLM9135 row: k column: 14
High quality sequence stop: 577.

FEATURES
source

1..579
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3964933"
/clone_lib="NCI_CGAP_Mam1"
/tissue_type="tumor, biopsy sample"
/dev_stage="10 months, virgin"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; site_1: SalI;
site_2: NotI; Cloned unidirectionally. Primer: Oligo dT
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

BASE COUNT
164 a 150 c 142 g 123 t

ORIGIN

Query Match 77.1%; Score 32.4; DB 10; Length 579;


```

JOURNAL      Unpublished (1999)
COMMENT      Other_ESTS: dafe8f06.xl
              Contact: Sandy Clifton, Ph.D.
              Washu Xenopus EST project, 1999
              Washington University School of Medicine
              4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
              Tel: 314 286 1800
              Fax: 314 286 1810
              Email: est@watson.wustl.edu
              cDNA Library Preparation: Life Technologies, Inc.
              CDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Washington University Genome Sequencing Center
              Clone distribution: Xenopus clones from this library are available
              through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
              High quality sequence stop: 416.
FEATURES     Location/Qualifiers
            source          1..423
                        /organism="Xenopus laevis"
                        /db_xref="taxon:8355"
                        /clone="IMAGE:4743706"
                        /clone_1ib="NICHD XGC Eye1"
                        /dev_stage="adult"
                        /lab_host="PH10B (phage-resistant)"
                        /note="Organ: eye; Vector: pCMV-Sport6. Site.1: Not;
                          Site.2: Salt; Cloned unidirectionally. Primer: Oligo dT.
                          Average insert size 2.3 kb. Constructed by life
                          technologies. Note: This is a Xenopus Gene Collection (XGC)
                          library."
BASE COUNT   139 a           86 c           137 g           61 t
ORIGIN
Query Match       50.0%; Score 25.2; DB 10; Length 423;
Best Local Similarity 78.9%; Pred. No. 1.le+02;
Matches          30; Conservative 0; Mismatches 8; Indels 0; Gaps 0
Ox               4 cctgctgttgcgtcgtcctccgcagcgacgaata 41
                ||| | | | | | | | | | | | | | | | | | | |
Db             201 CCAGCGGGGTGATGAGTGTCATCATGCGCIGACGAGTA 238

RESULT 14
LOCUS        BJO28868                               548 bp    mRNA    linear    EST 06-DEC-2001
DEFINITION   BJO28868 NIBB Mochii normalized Xenopus neurula library xenopus
ACCSSION     BJO28868                                NM010109 5', mRNA sequence.
VERSION      BJO28868.1 GI:17398582
KEYWORDS     EST.
SOURCE       African clawed frog.
ORGANISM     Xenopus laevis
REFERENCE    1 Eukaryota: Metazoa: Chordata: Craniata: Vertebrata; Euteleostomi;
AUTHORS      Amphibia: Batrachia: Anura: Mesobatrachia: Pipidoidea; Pipidae;
              Xenopodinae; Xenopus.
              1 (bases 1 to 548).
              Kiteyama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-I,T. and Kohara
              ,Y.
              Expressed genes in X. laevis embryo
              Unpublished (2001)
              Contact: Tadasu Shin-I
              Center For Genetic Resource Information
              National Institute of Genetics
              111 Yata, Mishima, Shizuoka 411-8540, Japan
              Tel: 81-559-81-6856
              Fax: 81-559-81-6855
              Email: tshinigenes.nig.ac.jp.
FEATURES     Location/Qualifiers
            1..548
                        /organism="Xenopus laevis"
                        /db_xref="taxon:8355"
                        /clone="XL010109"
                        /clone_1ib="NIBB Mochii normalized Xenopus neurula
                        library"

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 11, 2002, 03:10:28 ; Search time 1829.97 Seconds
(without alignments)
480.289 Million cell updates/sec

Title: US-09-786-214-11
Perfect score: 42
Sequence: 1 ctccctgctgtgtgtgtct.....ctccctgctgagcagagatcat 42

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: gb_ba:*
2: gb_hlg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_ph:*
7: gb_pl:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pac:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hlg_hum:*
31: em_hlg_inv:*
32: em_hlg_other:*
33: em_higo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Query	Score	Match	Length	DB	ID	Description
SUMMARIES							

1	42	100.0	361	9	HUMCSF1M1	M1038 Human macro
2	42	100.0	693	6	E03655	E03655 DNA sequence
3	42	100.0	821	6	I01437	I01437 Sequence 4
4	42	100.0	821	6	I09299	I09299 Sequence 3
5	42	100.0	931	9	HUMCSF1X	M76453 Homo sapien
6	42	100.0	1119	6	E12942	E12942 Human CDNA
7	42	100.0	1370	9	HSU22386	U22386 Human macro
8	42	100.0	1640	6	I08431	I08431 Sequence 4
9	42	100.0	1641	6	I00368	I00368 Sequence 1
10	42	100.0	1641	6	I01429	I01429 Sequence 1
11	42	100.0	1642	6	ARI05651	ARI05651 Sequence
12	42	100.0	1642	6	ARI10609	ARI10609 Sequence
13	42	100.0	1642	6	ARI141545	ARI141545 Sequence
14	42	100.0	1642	6	ARI142870	ARI142870 Sequence
15	42	100.0	1642	6	I01440	I01440 Sequence 7
16	42	100.0	1642	6	I08736	I08736 Sequence 2
17	42	100.0	1642	6	I09203	I09203 Sequence 1
18	42	100.0	1642	6	I09298	I09298 Sequence 1
19	42	100.0	1642	6	I27815	I27815 Sequence 3
20	42	100.0	1642	6	I28747	I28747 Sequence 3
21	42	100.0	1642	6	I50653	I50653 Sequence 3
22	42	100.0	1642	6	I67640	I67640 Sequence 3
23	42	100.0	1642	6	I71134	I71134 Sequence 3
24	42	100.0	1665	6	E12943	E12943 Human CDNA
25	42	100.0	1665	9	HUMMCSF	M64592 Human macro
26	42	100.0	1987	6	I02044	I02044 Sequence 1
27	42	100.0	1997	6	I03518	I03518 Sequence 1
28	42	100.0	2475	9	HUMMCSF1A	M27087 Homo sapien
29	42	100.0	2468	6	E01949	E01949 CDNA encodl
30	42	100.0	2468	6	E02182	E02182 DNA encodl
31	42	100.0	2488	23	E09614	E09614 Human CDNA
32	42	100.0	2500	6	E01950	E01950 CDNA encodl
33	42	100.0	2500	6	E02183	E02183 DNA encodl
34	42	100.0	2500	23	E09615	E09615 Human CDNA
35	42	100.0	2851	9	BC021117	BC021117 Homo sapl
36	42	100.0	3391	6	AX281643	AX281643 Sequence
37	42	100.0	3939	9	HUMCSDF1	M37435 Human macro
38	42	100.0	3981	6	I08586	I08586 Sequence 1
39	42	100.0	3981	6	I09309	I09309 Sequence 1
40	42	100.0	4021	6	E03219	E03219 DNA encodl
41	42	100.0	4021	6	E03235	E03235 CDNA encodl
42	41	97.6	814	6	I05804	I05804 Sequence 9
43	40	95.2	1664	6	I00370	I00370 Sequence 3
44	40	95.2	1844	6	I01431	I01431 Sequence 3
45	40	95.2	2302	6	ARI05650	ARI05650 Sequence

ALIGNMENTS

RESULT 1
HUMCSF1M1
LOCUS
DEFINITION HUMCSF1M1 361 bp DNA linear PRI 08-AUG-1995
Human macrophage-specific colony-stimulating factor (CSF-1) gene,
exon 1.
ACCESSION M1038
VERSION M1038.1 GI:181140
KEYWORDS alternative splicing; colony stimulating factor.
SEGMENT 1 of 3
SOURCE Human DNA and pancreatic tumor cell line MIA-PaCa-2, CDNA to mRNA,
clone pCCSF-17.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Kawasaki,E.S., Ladhner,M.B., Wang,A.M., Van Arsdell,J., Warren,M.K.,
Coyne,M.Y., Schweickart,V.L., Lee,M.-T., Wilson,K.J., Boosman,A.,
Stanley,E.R., Ralph,P. and Mark,D.F.
TITLE Molecular cloning of a complementary DNA encoding human
macrophage-specific colony-stimulating factor (CSF-1)
JOURNAL Science 230 (4723), 291-296 (1985)
MEDLINE 86018828
COMMENT Draft entry and printed copy of sequence for [1] kindly provided by

REFERENCE 2 (bases 1 to 1370)
AUTHORS Cerretti, D.P.
TITLE Direct Submission
JOURNAL Submitted (08-MAR-1995) Douglas P. Cerretti, Immunex Corporation

FEATURES
source 51 University St., Seattle, WA 98101, USA
Location/Qualifiers
1. .1370
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="1p13-p21"
45. .1361
/note="m-csf gamma precursor"
/codon_start=1
/product="macrophage-colony stimulating factor gamma precursor"
/protein_id="AA64849.1"
/db_xref="GI:758781"
/translation="MTAPGAAGRCPTTWLIGLILVCLASRSTPEVSPYCSHMIGSCHLQRIIDSDMERSCOLTEPEFVDOELKDPVCYLKAFILVQIMDMFERNTPNALIVQIQEELSLKSCPTKDYEDHDKACRTFETPELLEKYKNFNETKILLDKWNITSKNCNNSFAECSSQDVVTKPCNCLYKPAIPSSDPASVSHQPLAPSMAYAGLWDESGTESSLPGBOPHTVDPGSAKORPSTCQSEPEPTPYVKDSTIGSPQRPSEVGAENPQMEDILDSAMGTNVPPEASGEASEIIPVQGETELSPSPGSGSMQTEPARSPNPLASAPLPASAKGOOPADVTGHEROSESSSPLOAESVPHLLVPSVILVLIAGGLTFRRRRSHQEPQRADSPLEQEGSPITQDDQVELPV"
45. .140
141. .1358
/note="m-csf gamma"
/product="macrophage-colony stimulating factor gamma"
1182. .1250
/note="encodes transmembrane domain"
misc_feature
BASE COUNT 305 a 416 c 382 g 267 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 42; DB 9; Length 1370;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 95 CTCCTGCTGTGTGCTGCTGCTGCGAGCAGAGAT 136
|||||
1 CTCCTGCTGTGTGCTGCTGCTGCGAGCAGAGAT 42
|||||

RESULT 8
LOCUS 108431 1640 bp linear PAT 02-DEC-1994
DEFINITION Sequence 4 from Patent WO 8604607.
ACCESSION 108431
VERSION 108431.1 GI:588857
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1640)
AUTHORS Kawasaki,E.S., Ladner,M.B., Van Arsdel,J.N., Wang,A.M. and Ralph,P.
TITLE RECOMBINANT COLONY STIMULATING FACTOR-1
JOURNAL Patent: WO 8604607-A 4 14-AUG-1986;
FEATURES
source Location/Qualifiers
1. .1640
/organism="unknown"
BASE COUNT 356 a 480 c 402 g 402 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 42; DB 6; Length 1640;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 228 CTCCTGCTGTGTGCTGCTGCTGCGAGCAGAGAT 269
|||||
1 CTCCTGCTGTGTGCTGCTGCTGCGAGCAGAGAT 42
|||||

RESULT 9
LOCUS 108431 1640 bp linear PAT 02-DEC-1994
DEFINITION Sequence 4 from Patent WO 8604607.
ACCESSION 108431
VERSION 108431.1 GI:588857
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1640)
AUTHORS Kawasaki,E.S., Ladner,M.B., Van Arsdel,J.N., Wang,A.M. and Ralph,P.
TITLE RECOMBINANT COLONY STIMULATING FACTOR-1
JOURNAL Patent: WO 8604607-A 4 14-AUG-1986;
FEATURES
source Location/Qualifiers
1. .1640
/organism="unknown"
BASE COUNT 356 a 480 c 402 g 402 t
ORIGIN

100368
LOCUS 100368 1641 bp ss-DNA linear PAT 21-MAY-1993
DEFINITION Sequence 1 from Patent US 4929700.
ACCESSION 100368
VERSION 100368.1 GI:314019
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1641)
AUTHORS Halenbeck,R., Kothe,K., Cowgill,C. and Laird,W.J.
TITLE Production of purified, biologically active, bacterially produced recombinant human CSF-1
JOURNAL Patent: US 4929700-A 1 29-MAY-1990;
COMMENT Cetus Corporation;
Emeryville, CA
On Jul 30, 1993 this sequence version replaced gi:285550.
FEATURES
source Location/Qualifiers
1. .1641
/organism="unknown"
BASE COUNT 356 a 476 c 406 g 403 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 42; DB 6; Length 1641;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 229 CTCCTGCTGTGTGCTGCTGCTGCGAGCAGAGAT 270
|||||
1 CTCCTGCTGTGTGCTGCTGCTGCGAGCAGAGAT 42
|||||

RESULT 10
LOCUS 101429 1641 bp ss-DNA linear PAT 21-MAY-1993
DEFINITION Sequence 1 from Patent US 4847325.
ACCESSION 101429
VERSION 101429.1 GI:270181
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1641)
AUTHORS Shadle,P.J., Kothe,K.E., Moreland,M. and Katre,N.
TITLE Conjugation of polymer to colony stimulating factor-1
JOURNAL Patent: US 4847325-A 1 11-JUL-1989;
Cetus Corporation;
Emeryville, CA
FEATURES
source Location/Qualifiers
1. .1641
/organism="unknown"
BASE COUNT 356 a 479 c 403 g 403 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 42; DB 6; Length 1641;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 229 CTCCTGCTGTGTGCTGCTGCTGCGAGCAGAGAT 270
|||||
1 CTCCTGCTGTGTGCTGCTGCTGCGAGCAGAGAT 42
|||||

RESULT 11
LOCUS ARI05651 1642 bp DNA linear PAT 14-FEB-2001
DEFINITION Sequence 3 from patent US 6103224.
ACCESSION ARI05651
VERSION ARI05651.1 GI:12819716
KEYWORDS
SOURCE Unknown.

ORGANISM	Unknown.
REFERENCE	Unclassified.
AUTHORS	1 (bases 1 to 1642) Ladner,M.B., Van Arsdel,J.N., Martin,G.A., Kawasaki,E.S., Coyne,M.Yee, Halenbeck,R.F. and Kotbs,K.E. N.gradient.2 CSF-1 (short form) and carboxy truncated fragments thereof
TITLE	Patent: US 6103224-A 3 15-AUG-2000; Location/Qualifiers 1..1642
JOURNAL	
FEATURES	/organism="unknown"
SOURCE	
BASE COUNT	356 a 480 c 403 g 403 t
ORIGIN	
Query Match	100.0%; Score 42; DB 6; Length 1642;
Best Local Similarity	100.0%; Pred. No. 2.2e-06;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Oy 1 ctccctgctgttgcctgctcctcccgagacagagat 42	
Dn 229 CTCCTGCTGTGTGGTCTGTCTCTCGACGAGAT 270	
RESULT 12	
LOCUS	AR110609 1642 bp DNA linear PAT 14-FEB-2001
DEFINITION	Sequence 3 from patent US 6117422.
ACCESSION	AR110609
VERSION	AR110609.1 GI:12827423
KEYWORDS	.
SOURCE	Unknown.
ORGANISM	Unclassified.
REFERENCE	1 (bases 1 to 1642)
AUTHORS	Ladner,M.B., Noble,J.A., Martin,G.A., Kawasaki,E.S., Coyne,M.Yee, Halenbeck,R.F. and Kotbs,K.E. N.gradient.2-CSF-1(long form) and carboxy truncated fragments thereof
TITLE	Patent: US 6117422-A 3 12-SEP-2000; Location/Qualifiers 1..1642
JOURNAL	
FEATURES	/organism="unknown"
SOURCE	
BASE COUNT	356 a 480 c 403 g 403 t
ORIGIN	
Query Match	100.0%; Score 42; DB 6; Length 1642;
Best Local Similarity	100.0%; Pred. No. 2.2e-06;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Oy 1 ctccctgctgttgcctgctcctccggagacagagat 42	
Ddb 229 CTCCTGCTGTGTGGTCTGTCTCTCGACGAGAT 270	
RESULT 13	
LOCUS	AR141545 1642 bp DNA linear PAT 08-AUG-2001
DEFINITION	Sequence 3 from patent US 6146851.
ACCESSION	AR141545
VERSION	AR141545.1 GI:15101061
KEYWORDS	.
SOURCE	Unknown.
ORGANISM	Unclassified.
REFERENCE	1 (bases 1 to 1642)
AUTHORS	Ladner,M.B., Van Arsdel,J.N., Martin,G.A., Kawasaki,E.S., Coyne,M.Yee, Halenbeck,R.F. and Kotbs,K.E. DNA encoding NV2 (long form) and carboxy truncated fragments thereof
TITLE	Patent: US 6146851-A 3 14-NOV-2000; Location/Qualifiers
JOURNAL	
FEATURES	

[illegible]

Wed Sep 11 05:49:05 2002

us-09-786-214-11.rge

Page 6

	Matches	42;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Oy	1	ctccctgctgttgcgtcctcctcctgacgacgagat	42							
Db	229	ctccctgctgttgcgtcctcctcctgacgacgagat	270							

Search completed: September 11, 2002, 04:49:11
Job time: 5923 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 11, 2002, 03:52:28 : Search time 177.6 Seconds
(without alignments)
406.027 Million cell updates/sec

Title: US-09-786-214-11

Perfect score: 42
Sequence: 1 ctccctgctgtgtgtctc.....ctcctggcgagcagagat 42

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq_032802:*
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	100.0	42	21	AAZ99677
2	42	100.0	60	21	AAZ99675
3	42	100.0	78	21	AAZ99672
4	42	100.0	129	21	AAZ99758
5	42	100.0	660	20	AAZ80211
6	42	100.0	693	13	AAO24188
7	42	100.0	756	19	AAV33210
8	42	100.0	771	14	AAO38875
9	42	100.0	771	19	AAV33209

10	42	100.0	861	7	AAZ99677	First half of part
11	42	100.0	1119	10	AAZ99677	Human macrophage
12	42	100.0	1317	14	AAZ99677	Sequence of the M
13	42	100.0	1640	7	AAZ99677	Sequence encoding
14	42	100.0	1641	11	AAZ99677	Clone pCSF17 enco
15	42	100.0	1642	9	AAZ99677	Human CSF-1 CDNA
16	42	100.0	1642	9	AAZ99677	Sequence of a CDNA
17	42	100.0	1642	10	AAZ99677	CDNA of plasmid pC
18	42	100.0	1642	10	AAZ99677	Sequence encoding
19	42	100.0	1642	10	AAZ99677	Colony stimulating
20	42	100.0	1642	17	AAZ99677	Human short form C
21	42	100.0	1642	18	AAZ99677	Gene for human co
22	42	100.0	1642	18	AAZ99677	Truncated human co
23	42	100.0	1642	18	AAZ99677	Human colony stim
24	42	100.0	1642	22	AAZ99677	Human CSF-1 codin
25	42	100.0	1642	21	AAZ99677	Sequence of the M
26	42	100.0	1642	21	AAZ99677	Human colony stim
27	42	100.0	1642	21	AAZ99677	Human CSF-1 short
28	42	100.0	1642	22	AAZ99677	Human colony stim
29	42	100.0	1642	22	AAZ99677	Human CSF-1 codin
30	42	100.0	1642	22	AAZ99677	Sequence of the M
31	42	100.0	1645	14	AAZ99677	HindIII fragment o
32	42	100.0	1846	17	AAZ99677	Human CSF-1 genom
33	42	100.0	2045	10	AAZ99677	Human colony stim
34	42	100.0	2237	15	AAZ99677	Human MCSF CDNA
35	42	100.0	2488	10	AAZ99677	CDNA of lambda-cm5
36	42	100.0	2488	14	AAZ99677	Sequence of clone
37	42	100.0	2490	9	AAZ99677	Sequence encoding
38	42	100.0	2500	9	AAZ99677	Sequence encoding
39	42	100.0	2500	10	AAZ99677	Lambda-cm11 CDNA
40	42	100.0	2500	14	AAZ99677	Sequence of clone
41	42	100.0	3391	24	AAZ99677	Human DNA sequence
42	42	100.0	3981	8	AAZ99677	Macrophage-colony
43	42	100.0	3981	10	AAZ99677	Sequence of a macr
44	42	100.0	4021	12	AAZ99677	Monocyte macrophag
45	40	95.2	2037	10	AAZ99677	Clone 4 contg. pC

ALIGNMENTS

RESULT 1	AAZ99677	standard; DNA: 42 BP.
ID	AAZ99677	
XX	AAZ99677	
AC	AAZ99677	
XX	12-JUL-2000	(first entry)
DT	XX	
XX	XX	
DE	XX	DNA derived from macrophage colony stimulating gene alternative ORF.
KW	XX	tumour rejection antigen; macrophage colony stimulating gene;
KW	XX	macrophage-colony stimulating factor; antigen presenting cell;
KW	XX	human leukocyte antigen; CD8+ cytotoxic T lymphocyte; ss.
XX	XX	
OS	XX	Synthetic.
OS	XX	Homo sapiens.
XX	XX	
FH	XX	Key
FT	XX	Location/Qualifiers
FT	XX	1..42
FT	XX	/tag- a
PN	XX	WO200013699-A1.
XX	XX	
PD	XX	16-MAR-2000.
XX	XX	
XX	XX	03-SEP-1999; 99WO-US20344.
PF	XX	
XX	XX	Partial major macr
XX	XX	Human CSF-1 nucleo
PR	XX	04-SEP-1998; 98US-0099077.
XX	XX	
PA	XX	(LUDW-) LUDWIG INST CANCER RES.
XX	XX	
PI	XX	Probst-Keppler M, Van Den Eynde B, Boon-Falleur T;

```
XX WPI: 2000-256859/22.
DR P-PSDB; AAY84266.
PT Isolated polypeptide used to treat subjects having a disorder
PT characterized by expression of alternative open reading frame
PT macrophage-colony stimulating factor comprises 25 amino acid residue
PT sequence -
XX
PS Example 2: Page 65; 74pp; English.
XX
CC The present sequence encodes a peptide which is derived from a tumour
CC rejection antigen precursor encoded by an alternative open reading frame
CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC the alternative ORF of macrophage-colony stimulating factor, when
CC presented by an antigen presenting cell having a human leukocyte antigen
CC (HLA) class I molecule, effectively induce the activation and
CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC acids derived from the alternate ORF of macrophage-colony stimulating
CC factor are useful for enriching selectively a population of T lymphocytes
CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC characterized by expression of the polypeptide, and for identifying
CC functional variants and mimetics.
XX
SQ Sequence 42 BP; 4 A; 11 C; 13 G; 14 T; 0 other;

Query Match          100.0%; Score 42; DB 21; Length 42;
Best Local Similarity 100.0%; Pred. No. 2,2e-06;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctccctgctgtgtgtgtgtgtctctctctctgagcagagat 42
    |||||||
DB 1 ctccctgctgtgtgtgtgtgtgtctctctctctgagcagagat 42

RESULT 2
AAZ99675 ID AAZ99675 standard; DNA; 60 BP.
XX
AC AAZ99675;
XX
DT 12-JUL-2000 (first entry)
XX
DE Truncated alternate reading frame of macrophage colony stimulating gene.
XX
KW tumour rejection antigen; macrophage colony stimulating gene;
KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..60
FT /*tag= a
FT /product= "tumour rejection antigen"

WO200013699-A1.
XX
PD 16-MAR-2000.
XX
PF 03-SEP-1999; 99WO-US20344.
XX
PR 04-SEP-1998; 98US-0099077.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Probst-Keppler M, Van Den Eynde B, Boon-Falleur T;
XX
PT WPI: 2000-256859/22.
DR P-PSDB; AAY85265.
XX
PT Isolated polypeptide used to treat subjects having a disorder
```

```
PT characterized by expression of alternative open reading frame
PT macrophage-colony stimulating factor comprises 25 amino acid residue
PT sequence -
XX
PS Example 2: Page 64; 74pp; English.
XX
CC The present sequence encodes a truncated tumour rejection antigen
CC precursor, and represents a truncated alternative open reading frame
CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC the alternative ORF of macrophage-colony stimulating factor, when
CC presented by an antigen presenting cell having a human leukocyte antigen
CC (HLA) class I molecule, effectively induce the activation and
CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC acids derived from the alternate ORF of macrophage-colony stimulating
CC factor are useful for enriching selectively a population of T lymphocytes
CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC characterized by expression of the polypeptide, and for identifying
CC functional variants and mimetics.
XX
SQ Sequence 60 BP; 8 A; 15 C; 21 G; 16 T; 0 other;

Query Match          100.0%; Score 42; DB 21; Length 60;
Best Local Similarity 100.0%; Pred. No. 2,2e-06;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctccctgctgtgtgtgtgtgtgtctctctctctgagcagagat 42
    |||||||
DB 10 ctccctgctgtgtgtgtgtgtgtgtctctctctctgagcagagat 51

RESULT 3
AAZ99672 ID AAZ99672 standard; DNA; 78 BP.
XX
AC AAZ99672;
XX
DT 12-JUL-2000 (first entry)
XX
DE Alternate reading frame of macrophage colony stimulating gene.
XX
KW Renal cell carcinoma; antigen; cytotoxic T lymphocyte;
KW tumour rejection antigen; macrophage colony stimulating gene;
KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..78
FT /*tag= a
FT /product= "tumour rejection antigen"

WO200013699-A1.
XX
PD 16-MAR-2000.
XX
PF 03-SEP-1999; 99WO-US20344.
XX
PR 04-SEP-1998; 98US-0099077.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Probst-Keppler M, Van Den Eynde B, Boon-Falleur T;
XX
PT WPI: 2000-256859/22.
DR P-PSDB; AAY84264.
XX
PT Isolated polypeptide used to treat subjects having a disorder
PT characterized by expression of alternative open reading frame
PT macrophage-colony stimulating factor comprises 25 amino acid residue
PT sequence -
```

PS Example 2; Page 63; 74pp; English.
XX
CC The present sequence encodes a tumour rejection antigen precursor,
CC and represents an alternative open reading frame (ORF) of human
CC macrophage colony stimulating gene. Peptides derived from the
CC alternative ORF of macrophage-colony stimulating factor, when presented
CC by an antigen presenting cell having a human leukocyte antigen (HLA)
CC class I molecule, effectively induce the activation and proliferation
CC of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic acids
CC derived from the alternate ORF of macrophage-colony stimulating factor
CC are useful for enriching selectively a population of T lymphocytes
CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC characterized by expression of the polypeptide, and for identifying
CC functional variants and mimetics.
XX
SQ Sequence 78 BP; 11 A; 17 C; 29 G; 21 T; 0 other;

Query Match 100.0%; Score 42; DB 21; Length 78;
Best Local Similarity 100.0%; Pred. No. 2.5e-06;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctccctgctgtgtgtctgtctctctcctgagcagagat 42
DB 10 ctccctgctgtgtgtgtctgtctctcctgagcagagat 51

RESULT 4
AA29758
ID AA29758 standard; DNA; 129 BP.
XX
AC AA29758;
XX
DT 12-JUL-2000 (first entry)
XX
DE Partial major macrophage colony stimulating gene sequence.
XX
KW tumour rejection antigen; macrophage colony stimulating gene;
KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..129
FT /*tag= a
XX
XX
PN WO200013699-A1.
XX
PD 16-MAR-2000.
XX
PF 03-SEP-1999; 99WO-US20344.
XX
PR 04-SEP-1998; 98US-0099077.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Probst-Keppler M, Van Den Eynde B, Boon-Falleur T;
XX
DR WPI: 2000-256859/22.
DR P-PSDB: AAT69496.
XX
XX
XX Isolated polypeptide used to treat subjects having a disorder
XX characterized by expression of alternative open reading frame
XX macrophage-colony stimulating factor comprises 25 amino acid residue
XX sequence -
XX
XX Disclosure: Fig 4; 74pp; English.
XX
XX The present sequence encodes a portion of macrophage-colony stimulating
XX factor. The specification describes a tumour rejection antigen precursor,
XX which is encoded by an alternative open reading frame (ORF) of human
XX macrophage colony stimulating gene. Peptides derived from the

CC alternative ORF of macrophage-colony stimulating factor, when presented
CC by an antigen presenting cell having a human leukocyte antigen (HLA)
CC class I molecule, effectively induce the activation and proliferation
CC of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic acids
CC derived from the alternate ORF of macrophage-colony stimulating factor
CC are useful for enriching selectively a population of T lymphocytes
CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC characterized by expression of the polypeptide, and for identifying
CC functional variants and mimetics.
XX
SQ Sequence 129 BP; 18 A; 40 C; 44 G; 27 T; 0 other;

Query Match 100.0%; Score 42; DB 21; Length 129;
Best Local Similarity 100.0%; Pred. No. 2.7e-06;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctccctgctgtgtgtgtctgtctctcctgagcagagat 42
DB 51 ctccctgctgtgtgtgtgtctgtctctcctgagcagagat 92

RESULT 5
AA80211
ID AA80211 standard; DNA; 660 BP.
XX
AC AA80211;
XX
DT 20-AUG-1999 (first entry)
XX
DE Human CSF-1 nucleotide sequence.
XX
KW Human: M-CSF-1; colony stimulating factor; angiogenesis; tumour;
KW proliferative disorder; wound healing; cellular proliferative disease;
KW diabetic retinopathy; cancer; solid tumour; ss.
XX
OS Homo sapiens.
XX
PN WO929345-A1.
XX
PD 17-JUN-1999.
XX
PF 04-DEC-1998; 98WO-US25791.
XX
PR 05-DEC-1997; 97US-0067591.
XX
PA (LJOL-) LA JOLIA INST EXPERIMENTAL MEDICINE.
XX
PI Borgstrom P, Bourdon MA, Deryugina E, Rao PS;
XX
DR WPI: 1999-385494/32.
XX
PT Inhibition of angiogenesis by macrophage intervention, useful for
XX inhibiting tumour growth
XX
PS Disclosure: Page 28; 30pp; English.
XX
XX The present invention describes a method for the inhibition a
XX angiogenesis in a cell population in a mammal comprising inhibiting a
XX host cells angiogenic effect in the said mammal. The method can be
XX used to inhibit angiogenesis in tumours, wounds surrounding cells or
XX cells characteristic of a proliferative disorder. Therefore the method
XX is useful in the treatment of wound healing and cellular proliferative
XX diseases, e.g. diabetic retinopathy, and cancers, especially associated
XX with the presence of solid tumours. The method can be used in combination
XX with additional treatment, including surgery, radiation therapy and
XX chemotherapy. The present sequence represents a nucleotide sequence of
XX human colony stimulating factor (CSF-1) as given in the present
XX invention.
XX
SQ Sequence 660 BP; 154 A; 186 C; 182 G; 138 T; 0 other;


```

RESULT 8
AAQ38875
ID AAQ38875 standard; CDNA; 771 BP.
XX
XX
AC AAQ38875;
XX
DE 06-MAY-1993 (first entry)
XX
XX
DE Sequence of the M-CSF clone, M-CSF -alpha.
XX
XX
KW Macrophage; colony stimulating factor; immune system; proliferation;
KW differentiation; interferon gamma; TNF; splicing; ss.
XX
FH Key Location/Qualifiers
FT 1..771
FT /*cag- a
FT /note= "encodes M-CSF-alpha"
FT sig-peptide 1..96
FT /*cag- b
FT mat-peptide 97..771
FT /*cag- c
XX
XX
PN US5171675-A.
XX
PD 15-DEC-1992.
XX
XX
PE 28-JUL-1988; 88US-0226050.
XX
XX
PR 28-JUL-1988; 88US-0226050.
PR 08-AUG-1989; 89US-0391218.
XX
XX
PA (CERR/) CERRETTI D P.
XX
PI Anderson DM, Cerretti DP, Cosman D, Gallis BM, Tushinski RJ;
XX
DR WPI: 1993-008579/01.
DR P-PSDB; AAR30197.
XX
XX
PT Isolated DNA useful for determining biological role of M-CSF -
PT comprises nucleotide sequence encoding a functional biologically
PT active human macrophage colony stimulating factor gamma protein
XX
XX
PS Disclosure; Fig 2; 18pp; English.
XX
XX
CC The sequence represents M-CSF clone -alpha, which is a primary
CC translaation prod. of an alternative splicing event of M-CSF, and
CC lacks bases 545 to 1438 of M-CSF -beta. The sequence represents
CC a precursor to biologically active M-CSF which is important in the
CC regulation of immune responses, by its ability to potentiate the
CC proliferation and differentiation of macrophages from immature
CC haematopoietic progenitor cells, and induces effector functions of
CC mature macrophages including secretion of interferon gamma, TNF and
CC non-M-CSF colony stimulating activities.
CC See also AAQ33028 and AAQ38876.
XX
XX
SQ Sequence 771 BP; 183 A; 204 C; 216 G; 168 T; 0 other:

```

```

Query Match 100.0%; Score 42; DB 14; Length 771;
Best Local Similarity 100.0%; Pred. No. 3.6e-06;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 ctccctgctgtgtgtgtctgtctctcctgagcagagagat 42
    |||||||
DB 51 ctccctgctgtgtgtgtgtctgtctcctcctgagcagagagat 92

```

```

RESULT 9
AAV33209
ID AAV33209 standard; DNA; 771 BP.
XX
XX
AC AAV33209;

```

```

XX
XX
DE 07-DEC-1998 (first entry)
XX
XX
DE Macrophage colony stimulating factor-alpha DNA.
XX
XX
DE Macrophage colony stimulating factor-alpha; M-CSF; human;
KW gene therapy; cancer; Leishmania; ss.
XX
XX
OS Homo sapiens.
XX
XX
PN W09839449-A1.
XX
PD 11-SEP-1998.
XX
XX
PE 04-MAR-1998; 98NO-US04802.
XX
XX
PR 04-MAR-1997; 97US-0038583.
XX
XX
PA (CHIR ) CHIRON CORP.
XX
XX
PI Dworki V, Koths KE, Manning WC;
XX
XX
DR WPI: 1998-495850/42.
DR P-PSDB; AAW70328.
XX
XX
PT Treatment of diseased cells, e.g. cancer cells - by transforming
PT cells with gene delivery vehicle capable of expressing a macrophage
PT colony stimulating factor-alpha mutant, used for, e.g. treating
PT cancer
XX
XX
PS Example 7; Page 62-63; 78pp; English.
XX
XX
CC This DNA sequence codes for macrophage-colony stimulating factor
CC alpha (M-CSFa, see AAW70328). A claimed method for reducing a
CC population of diseased cells comprises transfecting or transducing
CC at least 1 from the population of diseased cells with a gene
CC delivery vehicle (GDV) capable of expressing a M-CSFa mutant. Such
CC mutants (see AAW70329) have a decreased capacity to be proteolytically
CC processed and released from a cell membrane. Also claimed is a
CC method for reducing a population of diseased cells by: (1)
CC administering a GDV which infects a population of diseased cells
CC and which expresses an M-CSFa polypeptide and a prodrug activator
CC polypeptide, and (11) administering a prodrug which is converted by
CC the prodrug activator polypeptide in the diseased cells to an agent
CC that is toxic to the diseased cells, where at least some cells in
CC the population of diseased cells are destroyed. Expression of the
CC M-CSFa mutant in the population of cells can provide an animal with
CC an increased immune response against the diseased cells. The
CC methods can be used for treating diseases such as cancer or a
CC disease manifesting a population of aberrant cells where the
CC population is created by infection with a pathogen, e.g. Leishmania.
XX
XX
SQ Sequence 771 BP; 183 A; 205 C; 216 G; 167 T; 0 other:

```

```

Query Match 100.0%; Score 42; DB 19; Length 771;
Best Local Similarity 100.0%; Pred. No. 3.6e-06;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 ctccctgctgtgtgtgtgtctgtctcctgagcagagagat 42
    |||||||
DB 51 ctccctgctgtgtgtgtgtgtctgtctcctcctgagcagagagat 92

```

```

RESULT 10
AAN60588
ID AAN60588 standard; CDNA; 861 BP.
XX
XX
AC AAN60588;

```

```

DE 29-JUL-1991 (first entry)

```

```

DE First half of partial sequence encoding human colony stimulating

```

```

DE factor (CSF).
XX
XX Immune system; enhancement; interferon production; monocyte; ss.
XX
XX Homo sapiens.
OS
XX
XX key location/Qualifiers
XX exon 218..340
XX /*tag= a
XX
XX M08604607-A.
XX
XX 14-AUG-1986.
XX
XX 03-FEB-1986; 86WO-US00238.
XX
XX 21-JAN-1986; 86US-0821068.
XX
XX 05-FEB-1985; 85US-0698359.
XX
XX 30-APR-1985; 85US-0728834.
XX
XX 14-JUN-1985; 85US-0744924.
XX
XX 18-JUL-1985; 85US-0756814.
XX
XX 20-JUN-1986; 86US-0876819.
XX
XX
XX (CETU ) CETUS CORP.
XX
XX Kawasaki ES, Ladner MB, VanArsdell JN, Wang AM, Ralph P;
XX Coyne MY, Warren MK;
XX
XX WPI; 1986-225459/34.
XX
XX P-PSDB; AAP61505.
XX
XX New recombinant colony stimulating factor-1 protein - is obtd. by
XX recombinant DNA procedures for use in enhancing immune system and
XX for treating monocyes to increase interferon prodn.
XX
XX
XX Disclosure; Fig 4a-b; 82pp; English.
XX
XX The CSF-1 protein claimed is esp. human CSF-1, or it may have an
XX equiv. AA sequence, with deletion or conservative substitution of
XX one or more AA, esp. between positions 158 and 224; 51 and 52 and/or
XX 191,192 and193; 15-20 and/or 75-84. The protein is esp. mCSF-1, or
XX 158 deleted CSF-1 or asp (59)-CSFL.
XX
XX
XX Sequence 861 BP; 167 A; 204 C; 251 G; 199 T; 40 other;
SQ

```

```

Query Match 100.0%; Score 42; DB 7; Length 861;
Best Local Similarity 100.0%; Pred. No. 3.7e-06;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 ctccctgctgtgtgtgtgtgtgtctctctgagcagagagat 42
    |||||
Db 229 ctccctgctgtgtgtgtgtgtgtctctctgagcagagagat 270

```

```

RESULT 11
AAN90515
ID AAN90515 standard; cDNA; 1119 BP.
XX
XX AAN90515;
XX
XX 27-NOV-1989 (first entry)
XX
XX Human macrophage colony stimulating factor precursor cDNA.
XX
XX Macrophage colony stimulating factor; human; recombinant;
XX leukopenia; bone marrow transplant; anti-cancer drug;
XX infectious diseases; AGR-ON cell line; human leukemic T cells;
XX precursor protein.
XX
XX Homo sapiens (human).
OS
XX
XX key Location/Qualifiers
XX

```

```

FT CDS 1..1116
FT /*tag= a
XX
XX EP328061-A.
XX
XX 16-AUG-1989.
XX
XX 08-FEB-1989; 89EP-0102138.
XX
XX 08-FEB-1988; 88JP-0062841.
XX
XX (SAKA ) OTSUKA PHARM KK.
XX
XX Hirato T, Nakai S, Hong YM, Kouno N, Hirai Y;
XX
XX WPI; 1989-235429/33.
XX
XX P-PSDB; AAP91031.
XX
XX Human recombinant colony-stimulating factors
XX - obtd. using AGR-ON cell line, used for treating leukopenia
XX and against infectious diseases.
XX
XX Disclosure; fig 3; 73pp; English.
XX
XX cDNA encoding biologically active recombinant human macrophage colony
XX stimulating factor precursor protein (AAP91031). This can be used
XX as a drug for prevention and cure of diseases involving leukopenia,
XX as an auxiliary agent for bone marrow transplants, as a drug to
XX prevent and cure infectious diseases, and as an anti-cancer drug.
XX
XX See AAN90514-N90522, AAP91032-P91043, AAP90123.
XX
XX
XX Sequence 1119 BP; 257 A; 339 C; 300 G; 223 T; 0 other;
SQ

```

```

Query Match 100.0%; Score 42; DB 10; Length 1119;
Best Local Similarity 100.0%; Pred. No. 3.8e-06;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 ctccctgctgtgtgtgtgtgtgtctctctgagcagagagat 42
    |||||
Db 51 ctccctgctgtgtgtgtgtgtgtctctctgagcagagagat 92

```

```

RESULT 12
AAQ38876
ID AAQ38876 standard; cDNA; 1317 BP.
XX
XX AAQ38876;
XX
XX 06-MAY-1993 (first entry)
XX
XX Sequence of the M-CSF clone, M-CSF -gamma.
XX
XX Macrophage; colony stimulating factor; immune system; proliferation;
XX differentiation; interferon-gamma; TNF; splicing; ss.
XX
XX
XX key Location/Qualifiers
XX CDS 1..1317
XX /*tag= a
XX /*note= "encodes M-CSF-gamma"
XX sig_peptide 1..96
XX mat_peptide 97..1317
XX /*tag= b
XX /*tag= c
XX
XX US5171675-A.
XX
XX 15-DEC-1992.
XX
XX 28-JUL-1988; 88US-0226050.
XX
XX 28-JUL-1988; 88US-0226050.
XX
XX 08-AUG-1989; 89US-0391218.
XX

```

XX	(CERR/) CERRETTI D P.
PA	
XX	Anderson DM, Cerretti DP, Cosman D, Gallis BM, Tushinski RJ;
PI	WPI; 1993-008579/01.
XX	P-PDSB; AAR30198.
DR	
XX	
PT	Isolated DNA useful for determining biological role of M-CSF -
PT	comprises nucleotide sequence encoding a functional biologically
PT	active human macrophage colony stimulating factor gamma protein
XX	
PS	Disclosure; Fig 2; 18pp; English.
XX	
CC	The sequence represents M-CSF clone -gamma, which is a primary
CC	translation prod. of an alternative splicing event of M-CSF, and
CC	lacks bases 1091 to 1438 of M-CSF -beta. The sequence represents
CC	a precursor to biologically active M-CSF which is important in the
CC	regulation of immune responses, by its ability to potentiate the
CC	proliferation and differentiation of macrophages from immature
CC	haematopoietic progenitor cells, and induces effector functions of
CC	mature macrophages including secretion of Interferon gamma, TNF and
CC	non-M-CSF colony stimulating activities.
CC	See also AAQ33028 and AAQ38875.
XX	
SQ	Sequence 1317 BP; 297 A; 395 C; 362 G; 263 T; 0 other:
	Query Match 100.0%; Score 42; DB 14; Length 1317;
	Best Local Similarity 100.0%; Pred. No. 4e-06;
	Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	1 cccctcgtctgctgtgcctcctcctggcgagcaggatgat 42
Db	51 ccgccgcgtgtgttgctgtcctcctggcgagcaggatga 92
RESULT 13	
AAN60590	
ID	AAN60590 standard; cDNA; 1640 BP.
XX	
AC	AAN60590;
XX	
DT	29-JUL-1991 (first entry)
XX	
DE	Sequence encoding mature colony stimulating factor 1 (mcsf-1).
XX	
KM	Immune system; enhancement; Interferon production; monocyte; ss.
XX	
OS	Homo sapiens.
XX	
FH	
FT	Key Location/Qualifiers
CDS	178..948
FT	/tag= a
XX	
PM	WO8604607-A.
XX	
PD	14-AUG-1986.
XX	
PF	03-FEB-1986; 86WO-US00238.
XX	
PR	21-JAN-1986; 86US-0821068.
PR	05-FEB-1985; 85US-0688339.
PR	30-APR-1985; 85US-0728834.
PR	14-JUN-1985; 85US-0744924.
PR	18-JUL-1985; 85US-0756814.
PR	20-JUN-1986; 86US-0876819.
XX	
PA	(CETU) CETUS CORP.
XX	
PI	Kawasaki ES, Ladner MB, Vanatsdell JN, Wang AM, Ralph P;
XX	Coyne WY, Warren MK;
XX	

DR WPI; 1986-225459/34.
DR P-PSDB; AAB06667.

PT New recombinant colony stimulating factor-1 protein - is obtd. by
XX recombinant DNA procedures for use in enhancing immune system and
PT for treating monocytes to increase interferon prodn.
XX

PS Disclosure; Fig 5; 82pp; English.
XX

CC The CSF-1 protein claimed is esp. human CSF-1, or it may have an
CC equiv. AA sequence, with deletion or conservative substitution of
CC one or more Aa, esp. between positions 158 and 224; 51 and 52 and/or
CC 191, 192 and 193; 15-20 and/or 75-84. The protein is esp. mCSF-1, or
CC 158 deleted CSF-1 or asp (59)-CSF1. AAN60590 encodes a 32 residue
CC putative signal sequence, which is presumably cleaved upon
CC secretion from mammalian cells.
XX

SQ Sequence 1640 BP; 356 A; 479 C; 403 G; 402 T; 0 other;

Query Match 100.0%; Score 42; DB 7; Length 1640;
Best Local Similarity 100.0%; Pred. No. 4.1e+05;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0

OY 1 ctccctgctgttgcgtcgtcctcgagcaggagat 42
|||
Db 228 ctccctgctgttgcgtcgtcctcgagcaggagat 269

RESULT 14
AAO04901
ID AAO04901 standard; cDNA: 1641 BP.
XX
AC AAO04901;
XX
DT 17-OCT-1990 (first entry)
XX
DE Clone pCSCF17 encoding a short form of human CSF1.
XX
KW Colony stimulating factor; CSF-1; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT 179..946
CDS /*tag= a

US4929700-A.
PN
XX 29-MAY-1990.
PD
XX 08-APR-1988; 88US-0173428.
PF
XX 16-APR-1987; 87US-0040174.
PR
XX 08-APR-1988; 88US-0173428.
XX
PA (CETU) CETUS CORP.
PI
PI Halenbeck R, Roths K, Cowgill C, Laird WJ;
WPI; 1990-192981/25.
DR
DR P-PSDB; AAR05486.
XX

PT Purified, biologically active colony stimulating factor-1 dimer -
PT from bacterially produced recombinant CSF-1 monomer by
XX solubilisation and refolding.
XX
XX Disclosure; ; pp; English.
PS
CC DNA useful in a method of producing active CSF-1 heterodimers from
CC the monomeric gene product, by unfolding, solubilising and
CC refolding the protein.
XX

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2002, 06:50:27 ; Search time 28.7 seconds
(without alignments)
18,888 Million cell updates/sec

Title: US-09-786-214-12
Perfect score: 72
Sequence: 1 LPAVGLSPGQER 14

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Swissprot_40:*
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	43	59.7	211 1 COBL_METJA	Q58917 methanococ
2	41	56.9	508 1 COBL_MYCTU	Q10677 mycobacteri
3	39	54.2	115 1 KV3I_HUMAN	P04433 homo sapien
4	39	54.2	115 1 KV5I_MOUSE	P01642 mus musculu
5	39	54.2	278 1 PURU_HAENI	Q03432 haemophilus
6	39	54.2	516 1 CAAD_DROME	Q9Y4T3 drosophila
7	39	54.2	597 1 NR4I_RAT	P22829 rattus norv
8	39	54.2	1402 1 Y197_MOUSE	Q920W3 mus musculu
9	39	54.2	1636 1 BUD3_YEAST	P23558 saccharomyc
10	38	52.8	390 1 COBL_MYCTU	Q10671 mycobacteri
11	38	52.8	429 1 RSL_LEULA	P50889 leucostococ
12	38	52.8	507 1 CATI_PICAN	P30263 pichia angu
13	38	52.8	699 1 EFG_HAENI	P43925 haemophilus
14	38	52.8	700 1 EFG_PASNU	P57938 pasteurella
15	37	51.4	280 1 PMXA_MOUSE	Q62066 mus musculu
16	37	51.4	281 1 PMXA_RAT	Q62066 mus musculu
17	37	51.4	446 1 COBL_ARCTU	O29534 a coblamini
18	37	51.4	446 1 ENO1_MAIZE	P26301 zea mays (m
19	37	51.4	557 1 PUR6_VITAC	P55195 vigna aconit
20	37	51.4	637 1 WGLI_CAUCR	Q91P66 caulobacter
21	37	51.4	658 1 VGI8_BP7A	P13332 bacteriophag
22	37	51.4	753 1 CKMA_BACUF	O32321 bacillus th
23	37	51.4	992 1 EVC_HUMAN	P57679 homo sapien
24	37	51.4	2269 1 RRP1_SVAL	P35341 simian vtru
25	36.5	50.7	374 1 RGSK_BOVIN	P73348 bos taurus
26	36	50.0	129 1 KV3H_HUMAN	P04207 homo sapien
27	36	50.0	158 1 RSD_ECOLI	P31690 escherichia
28	36	50.0	162 1 RSD_SALTY	Q91916 salmoneilla
29	36	50.0	243 1 SUWT_SYNP7	P42451 synecococc
30	36	50.0	280 1 PURU_ECOLI	P37051 escherichia
31	36	50.0	280 1 PURU_SHITL	P38480 shigella fl
32	36	50.0	342 1 HDPK_AZOVI	P31878 azotobacter
33	36	50.0	348 1 HOXV_AZOVI	P40597 azotobacter

ALIGNMENTS

34	36	50.0	359 1 ALP2_PEA	P46257 plasm sativ
35	36	50.0	359 1 ALP_CICAR	O65735 cicer ariet
36	36	50.0	422 1 TKSU_PYRKO	P58570 pyrococcus
37	36	50.0	471 1 CD36_BOVIN	P26201 bos taurus
38	36	50.0	598 1 NR4I_CANFA	P31866 canis famill
39	36	50.0	750 1 EUS_CHICK	P07916 gallus galli
40	36	50.0	844 1 HEXA_STRPN	P10564 streptococc
41	36	50.0	926 1 PTM4_HUMAN	P29074 homo sapien
42	36	50.0	1103 1 VG37_BPARI	O94085 bacteriophag
43	36	50.0	1173 1 ATC2_YEAST	P38829 saccharomyc
44	36	50.0	1365 1 YAK1_SCHPO	O09919 schizosach
45	36	50.0	3703 1 ABFL_HUMAN	Q15911 homo sapien

RESULT 1
COBL_METJA STANDARD; PRT; 211 AA.
ID COBL_METJA
AC 058917:
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable precorrin-6x C5,15-methyltransferase [decarboxylating]
DE (EC 2.1.1.132) (Precorrin-6 methyltransferase) (Precorrin-6x
DE methylase)
DE COBL OR M01522.
OS Methanococcus jannaschli.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weisman J.E., Fuhmann J.L., Nguyen D.,
RA Usterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Kleink H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschli";
RL Science 273:1058-1073(1996).
CC -!- FUNCTION: CATALYZES THE METHYLATION OF BOTH C-5 AND C-15 IN
CC PRECORRIN-6Y TO FORM PRECORRIN-8X (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: 2 S-adenosyl-L-methionine + precorrin-6y = 2
CC S-adenosyl-L-homocysteine + precorrin-8x + CO(2).
CC -!- PATHWAY: COBALAMIN BIOSYNTHESIS.
CC -!- SIMILARITY: TO S-TYRPHIMORIN CBIE, ALSO, LOW, TO OTHER
CC METHYLASES INVOLVED IN COBALAMIN BIOSYNTHESIS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL database
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: U67593; AAB99541.1; -
CC TIGR: M01522; -
CC InterPro: IPR000878; Corrin_porph_methyltnf.
CC Pfam: PF00590; TP_methylase; 1.
CC Cobalamin biosynthesis; Porphyrin biosynthesis; Transferase;
CC Methyltransferase; Complete proteome.
KW SEQUENCE 211 AA; 23805 MW; 279A1A2B14365510 CRC64;

Query Match 59.7%; Score 43; DB 1; Length 211;

Best Local Similarity 54.5%; Pred. No. 3.8;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 4 VYGLSPGDEY 14
1111111111
Db 4 IVGIGPDGDEY 14

RESULT 2

COBI_MYCTU STANDARD; PRT; 508 AA.

AC Q10677;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cobalamin biosynthesis protein COBI [includes: Precorrin-2 C20-methyltransferase (EC 2.1.1.130) (S-adenosyl-L-methionine--precorrin-2 methyltransferase) (SP2MT); Precorrin-3 methylase (EC 2.1.1.-)].
DE COBI OR RV2066 OR MT2126 OR MTCY49.05.
GN MYCOBACTERIUM TUBERCULOSIS.
OS Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=H37RV;
RX MEDLINE=98295987; Pubmed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R., Stalston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.";
RT Nature 393:537-544(1998).
RL [2]

RN SEQUENCE FROM N.A.
RP STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O., Peterson J., Deboy R., Dodson R., Gwin M.L., Holt D., Hickey E., Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L., Bacher A., Utterback T., Weidman J., Khouli H., Gill J., Mikula A., Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains";
RT Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: METHYLATES PRECORRIN-2 AT THE C-20 POSITION TO PRODUCE PRECORRIN-3A (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + precorrin-2 = S-adenosyl-L-homocysteine + precorrin-3A.
CC -1- PATHWAY: COBALAMIN BIOSYNTHESIS.
CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS SUMT, CYSG, CBIF/COBM AND CBIL/COBI.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL: Z73966; CA98214.1; -;
DR EMBL: AE007063; AAK46406.1; -;
DR TIGR: MT2126; -;
DR InterPro: IPR000878; Corrin_porph_methyltrnf.
DR InterPro: IPR003043; Uropor_methyltransf.
DR Pfam: PF00590; TP_methylase; 2.
DR PROSITE: PS00839; SUMT_1; 1.

DR PROSITE; PS00840; SUMT_2; 1.
KW Cobalamin biosynthesis; Porphyrin biosynthesis; Transferase;
KW Methyltransferase; Multifunctional enzyme; Complete proteome.
FT DOMAIN 1 243
FT DOMAIN 244 508
SQ SEQUENCE 508 AA; 53910 MW; 95AC06f022c4dc1 CRC64;

Query Match 56.9%; Score 41; DB 1; Length 508;
Best Local Similarity 58.3%; Pred. No. 20;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 AVVGLSPGDEY 14
1111111111
Db 250 AVVGLSPGDEY 261

RESULT 3

KV31_HUMAN STANDARD; PRT; 115 AA.

AC P04433;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig kappa chain V-III region Vg precursor (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=85087932; Pubmed=6440122;
RA Pech M., Zachau H.G.;
RT "Immunoglobulin genes of different subgroups are interdigitated within the Vκ locus".
RL Nucleic Acids Res. 12:9229-9236(1984).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL: X01668; -; NOT_ANNOTATED_CDS.
DR PIR: A01900; K3HUVG.
DR HSSP: P01789; IMCP.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003596; Ig_V.
DR Pfam: PF00047; Ig_1.
DR SMART: SM00406; IgV_1.
KW Immunoglobulin V region; Signal.

FT SIGNAL 1 20
FT CHAIN 21 >115
FT DOMAIN 21 43
FT DOMAIN 44 54
FT DOMAIN 55 69
FT DOMAIN 70 76
FT DOMAIN 77 108
FT DOMAIN 109 115
FT DISULFID 43 108
FT NON_TER 115 115
SQ SEQUENCE 115 AA; 12575 MW; 2DE47CDA3A17D555 CRC64;

Query Match 54.2%; Score 39; DB 1; Length 115;
Best Local Similarity 63.6%; Pred. No. 9.9;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 PAVVGLSPGDEY 12
1111111111
Db 28 PATLSLSPGR 38


```

RESULT 4
KVS1_MOUSE STANDARD: PRT: 115 AA.
AC P01642;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig kappa chain V-V region L7 precursor (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=81220975; PubMed=6264318;
RA Pech M., Hochli J., Schnell H., Zachau H.G.;
RT "Differences between germ-line and rearranged immunoglobulin V kappa
RT coding sequences suggest a localized mutation mechanism.";
RL Nature 291:668-670(1981).
CC -1- MISCELLANEOUS: THERE APPEAR TO BE TWO POSSIBLE SPICE JUNCTIONS AT
CC THE 3' END OF THE INTRON. THE ALTERNATE WOULD CODE FOR A PROTEIN
CC LACKING RESIDUES 17-19.
DR PIR: A01925; KWSL7.
DR InterPro: IPR003006; IG_MHC.
DR InterPro: IPR003596; IG_V.
DR Pfam: PF00047; Ig_1.
DR SMART: SM00406; IgV_1.
KW Immunoglobulin V region; Signal.
FT SIGNAL 1 20
FT CHAIN 21 >115 IG KAPPA CHAIN V-V REGION L7.
FT DOMAIN 21 43 FRAMEWORK-1.
FT DOMAIN 44 54 COMPLEMENTARITY-DETERMINING-1.
FT DOMAIN 55 69 FRAMEWORK-2.
FT DOMAIN 70 76 COMPLEMENTARITY-DETERMINING-2.
FT DOMAIN 77 108 FRAMEWORK-3.
FT DOMAIN 109 >115 COMPLEMENTARITY-DETERMINING-3.
FT DISULFID 43 108 BY SIMILARITY.
FT NON_TER 115 115
SQ SEQUENCE 115 AA; 12615 MW; C17BEC758C577E00 CRC64;

Query Match 54.2%; Score 39; DB 1; Length 115;
Best Local Similarity 54.5%; Pred. No. 9.9;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 2 PAVVGLSPGEQ 12
DB 28 PALLSVSPGER 38
||:|||||
|:::|||||

RESULT 5
PURU_HAETIN STANDARD: PRT: 278 AA.
ID Q03432;
AC Q03432;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Formyltetrahydrofolate deformylase (EC 3.5.1.10) (Formyl-FH(4)
DE hydrolase).
GN PURU OR H11588.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_Taxid=727;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,

```

```

RA Scott J.D., Shirley R., Liu L.-I., Glodok A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uitterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Genhm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd. ";
RL Science 269:496-512(1995).
RN [2]
RP SEQUENCE OF 64-278 FROM N.A.
RC STRAIN=RM 7004;
RX MEDLINE=93328119; PubMed=8335255;
RA Maskell D.J.;
RT "Cloning and sequencing of the Haemophilus influenzae araC gene.";
RL Gene 129:155-156(1993).
CC -1- FUNCTION: PRODUCES FORMATE FROM FORMYL-TETRAHYDROFOLATE. PROVIDES
CC THE MAJOR SOURCE OF FORMATE FOR THE PURT-DEPENDENT SYNTHESIS OF
CC 5'-PHOSPHORIBOSYL-N-FORMYLGLYCINAMIDE (FGAR) DURING AEROBIC
CC GROWTH. HAS A ROLE IN REGULATING THE ONE-CARBON POOL
CC (BY SIMILARITY)
CC -1- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + H(2)O = formate +
CC tetrahydrofolate.
CC -1- ENZYME REGULATION: ACTIVATED BY METHIONINE, INHIBITED BY GLYCINE
CC (BY SIMILARITY).
CC -1- PATHWAY: DE NOVO PURINE BIOSYNTHESIS.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY)
CC -1- SIMILARITY: SOME, TO GAR TRANSFORMYLASE (PURN).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC -----
DR EMBL: U32833; AAC23236.1; -.
DR EMBL: L04686; AAA24942.1; -.
DR PIR: P06066; P06066.
DR HSSP: P08179; IGRC.
DR TIGR: H11588; -.
DR InterPro: IPR002912; ACT.
DR InterPro: IPR002376; formyl_transf.
DR Pfam: PF01842; ACT; 1.
DR Pfam: PF00551; formyl_transf; 1.
KW Purine biosynthesis; Hydrolase; One-carbon metabolism;
KW Complete Proteome.
FT ACT_SITE 223 BY SIMILARITY.
FT CONFLICT 115 117 VIG -> RNR (IN REF. 2).
FT CONFLICT 138 140 HEN -> PK (IN REF. 2).
FT CONFLICT 205 205 K -> E (IN REF. 2).
SQ SEQUENCE 278 AA; 32173 MW; 7F375AB3C422EC4B CRC64;

Query Match 54.2%; Score 39; DB 1; Length 278;
Best Local Similarity 50.0%; Pred. No. 24;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 LPAVGLSPGEQY 14
DB 191 LPAFIGAKPYQOAY 204
|||:| | | |
|:::| | | |

RESULT 6
C4AD_DROME STANDARD: PRT: 516 AA.
ID C4AD_DROME
AC Q9V4T3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable cytochrome P450 4d1 (EC 1.14.-.-) (CYP4VAD1).
GN CYP4AD1 OR CG2110.

```

OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NC NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkov D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
 RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Crawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kohira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheefer F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Sytkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weissstock G.M., Weissbach J.,
 RA Williams S.M., Wodner T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 CC -1- FUNCTION: MAY BE INVOLVED IN THE METABOLISM OF INSECT HORMONES AND
 CC IN THE BREAKDOWN OF SYNTHETIC INSECTICIDES (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: RN + reduced flavoprotein + O(2) -> ROH +
 CC oxidized flavoprotein + H(2)O.
 CC -1- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum
 CC (Potential).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AEO03837; AAF59092.1;
 DR FLYBase: FBgn0033292; Cyp4d1.
 DR InterPro: IPR001128; Cyt_P450.
 DR Pfam: PF00067; P450.1.
 DR PRINTS: PR00385; P450.
 DR PROSITE: PS00086; CYTOCHROME_P450; 1.
 KW Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
 KW Endoplasmic reticulum; Hypothetical protein.
 FT BINDING 445 445 HEME (BY SIMILARITY).
 SQ SEQUENCE 516 AA; 58870 MW; 648E2A2492AF58C7 CRC64;

Query Match 54.2%; Score 39; DB 1; Length 516;
 Best Local Similarity 57.1%; Pred. No. 44;
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 QY 1 LPVAVGLSPGEQY 14
 Db 469 LPVAVGLPPGINDH 482
 ||| ||| :
 : : :
 RESULT 7
 NR41_RAT ID NR41_RAT STANDARD; PRT; 597 AA.
 AC P22829;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Orphan nuclear receptor HMR (Nerve growth factor induced protein I-B)
 DE (NGFI-B) (NMR77).
 GN NR4A1 OR HMR OR NGFI-B.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 NC NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90166506; PubMed=3272167;
 RA Milbrandt J.;
 RT "Nerve growth factor induces a gene homologous to the glucocorticoid
 RT receptor gene.";
 RL Neuron 1:183-188(1988).
 RL [2]
 RP CHARACTERIZATION.
 RX MEDLINE=93361012; PubMed=8395013;
 RA Wilson T.E., Fahrner T.J., Milbrandt J.;
 RT "The orphan receptors NGFI-B and steroidogenic factor 1 establish
 RT monomer binding as a third paradigm of nuclear receptor-DNA
 RT interaction.";
 RL Mol. Cell. Biol. 13:5794-5804(1993).
 RL [3]
 RP DNA BINDING MOTIFS.
 RX MEDLINE=92229411; PubMed=1314418;
 RA Hirata Y., Kiuchi K., Chen H.-C., Milbrandt J., Garoff G.;
 RT "The phosphorylation and DNA binding of the DNA-binding domain of the
 RT orphan nuclear receptor NGFI-B.";
 RL J. Biol. Chem. 268:24808-24812(1993).
 CC -1- FUNCTION: PROBABLE NUCLEAR RECEPTOR. MAY ACT CONCOMITANTLY WITH
 CC NR4I IN REGULATING THE EXPRESSION OF DELAYED-EARLY GENES DURING
 CC LIVER REGENERATION. RECOGNIZES THE SINGLE HALF-SITE 5'-AAAGGTCA-
 CC 3'.
 CC -1- SUBUNIT: BINDS DNA AS A MONOMER.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE LUNG, BRAIN, SUPERIOR
 CC CEREBRAL GANGLIA. HIGH LEVELS SEEN IN THE ADRENAL TISSUE.
 CC -1- INDUCTION: BY NERVE GROWTH FACTOR AND DURING LIVER REGENERATION.
 CC -1- PTM: PHOSPHORYLATION OF SER-350 RESULTS IN DECREASE IN NBRE
 CC BINDING WHILE PHOSPHORYLATION OF SER-340 HAS LITTLE EFFECT ON IT.
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
 CC NR4 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

entitles requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: U17254; AAA56770.1; ALT_INIT.
 DR PIR: J00623; J00623.
 DR HSSP: P19793; 2NLE.
 DR TRANSFAC: T00619; -.
 DR InterPro: IPR000536; Hormone_rec_11g.
 DR InterPro: IPR001723; Strdhormone_receptor.
 DR InterPro: IPR001628; zf-C4.
 DR Pfam: PF00104; hormone_rec; 1.
 DR Pfam: PF00105; zf-C4; 1.
 DR PRINTS: PR00398; STRDHORMONER.
 DR PRINTS: PR00047; STROIDFINGER.
 DR SMART: SM00430; HOLI; 1.
 DR SMART: SM00399; znf_C4; 1.
 DR PROSITE: PS00031; NUCLEAR_RECEPTOR; 1.
 DR Receptor: Transcription regulation; DNA-binding; Nuclear protein; Zinc-finger; Phosphorylation.
 KM Zinc-finger: Phosphorylation.
 FT ZN_FING 266 331 NUCLEAR RECEPTOR-TYPE.
 FT ZN_FING 266 286 C4-TYPE.
 FT ZN_FING 302 326 C4-TYPE.
 FT DOMAIN 408 458 LIGAND-BINDING (POTENTIAL).
 FT MOD_RES 340 340 PHOSPHORYLATION (BY PKA).
 FT MOD_RES 350 350 PHOSPHORYLATION (BY PKA).
 FT MUTAGEN 340 340 S->A: LOSS OF PHOSPHORYLATION.
 FT MUTAGEN 350 350 S->A: LOSS OF PHOSPHORYLATION.
 FT MUTAGEN 345 345 R->K: DECREASED NBRE BINDING.
 FT MUTAGEN 348 348 L->V: ALMOST COMPLETE LOSS OF NBRE BINDING.
 FT SEQUENCE 597 AA; 64281 MW; 9CFA987112337E53 CRC64;

Query Match 54.2%; Score 39; DB 1; Length 597;
 Best Local Similarity 46.2%; Pred. No. 51;
 Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 1 LPAVGLSPGEOE 13
 DB 423 IPEFIELSPGDDO 435

RESULT 8
 Y197_MOUSE STANDARD: PRT; 1402 AA.
 ID Y197_MOUSE
 AC 0920M3;
 DT 16-OCT-2001 (Rel. 40; Created)
 DT 16-OCT-2001 (Rel. 40; Last sequence update)
 DE Protein KIAA0197 (GTL-13).
 GN KIAA0197 OR GTL-13.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 ON NCBI_TaxID=10090;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/SVT;
 RA Van de Putte T., Cozljensen M., Dewulf N., Tydzanowski P., Lomoy O., Huybrede D.;
 RA "Mus musculus mRNA for gtl-13 (gene trap locus-13), similar to human KIAA0197 gene (D83781), complete cds."
 RT Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
 RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR MGD: MGI:1926227; Gtl1-13.
 SQ SEQUENCE 1402 AA; 158230 MW; 3BF5D9F057D28772 CRC64;

Query Match 54.2%; Score 39; DB 1; Length 1402;
 Best Local Similarity 70.0%; Pred. No. 1.2e+02;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 LPAVGLSPG 10
 DB 170 IPEVPLSPG 179

RESULT 9
 BUD3_YEAST STANDARD: PRT; 1636 AA.
 ID BUD3_YEAST
 AC P25558; P25557; P87007;
 DT 01-MAY-1992 (Rel. 22; Created)
 DT 01-NOV-1995 (Rel. 32; Last sequence update)
 DT 16-OCT-2001 (Rel. 40; Last annotation update)
 DE Bud site selection protein BUD3.
 GN BUD3 OR YCL014W OR YCL114W/YCL11W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 ON NCBI_TaxID=4932;
 RX [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95247824; PubMed=7730410;
 RA Chant J., Mische M., Mitchell E., Herskowitz I., Pringle J.R.;
 RT "Role of Bud3 in producing the axial budding pattern of yeast."
 RL J. Cell Biol. 129:767-778(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Oliver S.G., Anwar R., Brown A., Gent M.E., Indge K.J., James C.M., Staceva L.I.;
 RL Submitted (MAR-1992) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP REVISIONS.
 RA Gromadka R.;
 RL Submitted (JAN-1996) to the EMBL/Genbank/DBJ databases.
 RN [4]
 RP REVISIONS.
 RA Mewes H.-W.;
 RL Submitted (JUN-2001) to the EMBL/Genbank/DBJ databases.
 CC -!- FUNCTION: CO-ASSEMBLES WITH BUD4 AT BUD SITES. BUD4 AND BUD3 MAY COOPERATE TO RECOGNIZE A SPATIAL LANDMARK (THE NECK FILAMENTS) DURING MIOSIS AND THEY SUBSEQUENTLY BECOME A LANDMARK FOR ESTABLISHING THE AXIAL BUDDING PATTERN IN G1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: U17580; AAA6315.1; -.
 DR EMBL: X59720; CAA42346.2; -.
 DR PIR: S19341; S19341.
 DR PIR: S19340; S19340.
 DR SGD: S000520; BUD3.
 DR InterPro: IPR000219; RhogEF.
 DR SMART: SM00325; RhogEF; 1.
 SQ SEQUENCE 1636 AA; 184717 MW; 9E4E6BA5C3A3F69 CRC64;

Query Match 54.2%; Score 39; DB 1; Length 1636;
 Best Local Similarity 58.3%; Pred. No. 1.4e+02;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 2 PAVGLSPGEOE 13

```

Db      112 PATENTSPDOE 123

      11: 111:11
RESULT 10
COBL_MCTU STANDARD: PRT: 390 AA.
ID      COBL_MCTU
AC      Q10671;
DT      15-JUL-1998 (Rel. 36, Created)
DT      15-DEC-1998 (Rel. 37, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Precorrin-6y C5,15-methyltransferase [decaoxylating] (EC 2.1.1.132)
GN      COBL OR RV2072C OR MT2132 OR MTCY49.11C.
OS      Mycobacterium tuberculosis.
OC      Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC      Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX      NCBI_TaxID=1773;
[1]
RN      SEQUENCE FROM N.A.
RP      STRAIN-H37RV;
RC      MEDLINE=98295987; PubMed=9634230;
RA      Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA      Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA      Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA      Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA      Horsley T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA      Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA      Ruter S., Seeger K., Skelton S., Squares S., Squares R.,
RA      Sulston J.E., Taylor K., Whitehead S., Barrett B.G.,
RT      "Deciphering the biology of Mycobacterium tuberculosis from the
RT      complete genome sequence."
RL      Nature 393:537-544(1998).
[2]
RN      SEQUENCE FROM N.A.
RP      STRAIN-CDC 1551 / Oshkosh;
RC      Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA      Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA      Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA      Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA      Bishai W.;
RT      "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT      laboratory strains."
RL      Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC      -I- FUNCTION: CATALYZES THE METHYLATION OF BOTH C-5 AND C-15 IN
CC      PRECORRIN-6Y TO FORM PRECORRIN-8X.
CC      -I- CATALYTIC ACTIVITY: 2 S-adenosyl-L-methionine + precorrin-6Y = 2
CC      S-adenosyl-L-homocysteine + precorrin-8X + Co(2).
CC      -I- PATHWAY: COBALAMIN BIOSYNTHESIS.
CC      -I- SIMILARITY: TO S-TYRPHIMURION CBIE; ALSO, LOW, TO OTHER
CC      METHYLASES INVOLVED IN COBALAMIN BIOSYNTHESIS.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; Z73966; CAA98325.1; -
DR      EMBL; AE007063; AAK46412.1; -
DR      TIGR; MT2132; -
DR      Tuberculin; RV2072c; -
DR      InterPro; IPR000878; Corrin_porph_methyltrnf.
DR      InterPro; IPR000051; SAM_bind.
DR      Pfam; PF00590; TP_methylase; 1.
KW      Cobalamin biosynthesis; Porphyrin biosynthesis; Transferase;
KW      Methyltransferase; Complete proteome.
FT      CONFLICT 205 L->P (IN REF. 2).
FT      CONFLICT 327 327 D->H (IN REF. 2).
SQ      SEQUENCE 390 AA; 41854 MW; F842EFF7562F21F3 CMC64;

```

```

Query Match      52.8%; Score 38; DB 1; Length 390;
Best Local Similarity 88.9%; Pred. No. 49;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      1 LPAVGLSP 9
      11111111
Db      55 LPAVGLSP 63

RESULT 11
RSL_LEU1A
ID      RSL_LEU1A STANDARD: PRT: 429 AA.
AC      P50889; P71450;
DT      01-OCT-1996 (Rel. 34, Created)
DT      15-JUL-1999 (Rel. 38, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      40S ribosomal protein S1.
GN      RPS1.
OS      Leuconostoc lactis.
OC      Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC      Leuconostoc.
OX      NCBI_TaxID=1246;
[1]
RN      SEQUENCE FROM N.A.
RP      MEDLINE=97186703; PubMed=9034319;
RA      Yamit-Hezi A., Levy Z., Neuman S.,
RT      "A leuconostoc lactis protein with homology to ribosomal protein S1
RT      shares common epitopes and common DNA binding properties with a
RT      mammalian DNA binding nuclear factor."
RL      Gene 185:99-103(1997).
[2]
RN      SEQUENCE OF 24-429 FROM N.A.
RP      MEDLINE=95237615; PubMed=7721096;
RA      Eklund E.A., Lee S.W., Skalniak D.G.;
RT      "Cloning of a cDNA encoding a human DNA-binding protein similar to
RT      ribosomal protein S1."
RL      Gene 155:231-235(1995).
[3]
RN      SEQUENCE OF 78-429 FROM N.A.
RP      MEDLINE=96164600; PubMed=8568274;
RA      Tsurzaka K., Leu A.K., Frank M.B., Movafagh B.F., Kosec M.,
RA      Winkler T.H., Kalden J.R., Reichlin M.;
RT      "Lupus autoantibodies to double-stranded DNA cross-react with
RT      ribosomal protein S1."
RL      J. Immunol. 156:1668-1675(1996).
CC      -I- FUNCTION: EXHIBITS PREFERENTIAL BINDING TO SINGLE-STRANDED AND
CC      DOUBLE-STRANDED DNA AND A LOW BINDING AFFINITY FOR RNA.
CC      -I- SIMILARITY: BELONGS TO THE S1P FAMILY OF RIBOSOMAL PROTEINS.
CC      -I- SIMILARITY: CONAINS 4 S1 MOTIF DOMAINS.
CC      -I- CAUTION: WAS ORIGINALLY (REF.2 AND REF.3) THOUGHT TO ORIGINATE
CC      FROM HUMAN BUT IS MOST PROBABLY THE RESULT OF A CDNA LIBRARY
CC      CONTAMINATION BY L.LACTIS.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; U24086; AAB08978.1; -
DR      EMBL; U05589; AAA7669.1; -
DR      EMBL; U27517; AAA97575.1; -
DR      HSSP; P05055; ISRO.
DR      InterPro; IPR000110; Ribosomal_S1.
DR      InterPro; IPR003029; S1.
DR      Pfam; PF00575; S1; 4.
DR      PRINTS; PR00681; RIBOSOMALS1.
DR      SMART; SM00316; S1; 4.
DR      PROSITE; PS50126; S1; 4.
KW      Ribosomal protein; Repeat; RNA-binding.

```

FT DOMAIN 55 128 S1 MOTIF 1.
 FT DOMAIN 144 211 S1 MOTIF 2.
 FT DOMAIN 231 299 S1 MOTIF 3.
 FT DOMAIN 316 385 S1 MOTIF 4.
 FT CONFLICT 24 24 S -> G (IN REF. 2).
 FT CONFLICT 122 122 A -> S (IN REF. 3).
 FT CONFLICT 217 217 L -> R (IN REF. 2 AND 3).
 SQ SEQUENCE 429 AA: 46386 MW: 92AC82605F39D9FC CRC64:

Query Match Best Local Similarity 52.8%; Score 38; DB 1; Length 429;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 3 AVGLSPGSEQ 12
 Db 71 AVGLSTGEE 80

RESULT 12
 CATA_PICAN STANDARD; PRT; 507 AA.
 AC P30263;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Catalase (EC 1.11.1.6).
 GN PXP9 OR PXP-9.
 OS Pichia angusta (Yeast) (Hansenula polymorpha).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Pichia.
 OX NCBI_TaxID=4905;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 34438;
 RX MEDLINE=92299073; PubMed=1607006;
 RA Dildion T., Roggenkamp R.O.;
 RT Targeting signal of the peroxisomal catalase in the methylotrophic
 RT yeast Hansenula polymorpha";
 RL FEBS Lett. 303:113-116(1992).
 CC -1- FUNCTION: OCCURS IN ALMOST ALL AEROBICALLY RESPIRING ORGANISMS AND
 CC SERVES TO PROTECT CELLS FROM THE TOXIC EFFECTS OF HYDROGEN
 CC PEROXIDE.
 CC -1- CATALYTIC ACTIVITY: 2 H(2)O(2) - O(2) + 2 H(2)O.
 CC -1- COFACTOR: HEME GROUP.
 CC -1- SUBUNIT: HOMOTETRAMER.
 CC -1- SUBCELLULAR LOCATION: Peroxisomal.
 CC -1- SIMILARITY: BELONGS TO THE CATALASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: X56501; CAA39856.1; -
 CC PIR: S23422; S23422.
 CC HSSP: P15202; 1A4E.
 DR InterPro: IPR002226; Catalase.
 DR Pfam: PF00199; catalase; 1.
 DR PRINTS: PR00067; CATALASE.
 DR PRODOM: PD000510; Catalase; 1.
 DR PROSITE: PS00342; MICRODIOLDES_CTER; 1.
 DR PROSITE: PS00437; CATALASE_1; 1.
 DR PROSITE: PS00438; CATALASE_2; 1.
 KW Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide;
 KM Peroxisome.
 FT ACT_SITE 65 65 BY SIMILARITY.
 FT ACT_SITE 138 138 BY SIMILARITY.
 FT BINDING 348 348 PROXIMAL HEME LIGAND (BY SIMILARITY).
 FT SITE 505 507 MICROBODY TARGETING SIGNAL (POTENTIAL).

SQ SEQUENCE 507 AA: 57849 MW: 3536ED0A49539CC3 CRC64:

Query Match Best Local Similarity 52.8%; Score 38; DB 1; Length 507;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 4 VGLSPGSEQ 13
 Db 445 VGLRTPGSEQ 454

RESULT 13
 EFG_HAEIN STANDARD; PRT; 699 AA.
 AC P43925;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Elongation factor G (EF-G).
 GN FUSA OR FUS OR HI0579.
 GN Haemophilus influenzae.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RD / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shiley R., Liu L.-I., Glodok A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fitchman J.L., Fuhmann J.L., Geoghegan N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT Whole-genome random sequencing and assembly of Haemophilus
 RT influenzae Rd. ";
 RL Science 269:496-512(1995).
 CC -1- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT TRANSLOCATION
 CC OF THE NASCENT PROTEIN CHAIN FROM THE A-SITE TO THE P-SITE OF THE
 CC RIBOSOME.
 CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
 CC -1- EF-G/EF-2 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: U37239; AAC2237.1; -
 CC HSSP: P13551; 1FNM.
 DR TIGR: HI0579; -
 DR InterPro: IPR000640; EFG_C.
 DR InterPro: IPR000795; GTP_EFTU.
 DR InterPro: IPR004161; GTP_EFTU_D2.
 DR Pfam: PF00679; EFG_C; 1.
 DR Pfam: PF00009; GTP_EFTU; 1.
 DR Pfam: PF03144; GTP_EFTU_D2; 1.
 DR PRINTS: PR00315; ELONGINFCT.
 DR PROSITE: PS00301; EFACOR_GTP; 1.
 KW Elongation factor; Protein biosynthesis; GTP-binding;
 KM Complete proteome.
 FT INIT_MET 0 0 BY SIMILARITY.
 FT NP_BIND 16 23 GTP (BY SIMILARITY).
 FT NP_BIND 87 91 GTP (BY SIMILARITY).
 FT NP_BIND 141 144 GTP (BY SIMILARITY).

SEQ SEQUENCE 699 AA; 77132 MW; FBBAD639C0F62801 CRC64;

Query Match 52.8%; Score 38; DB 1; Length 699;
Best Local Similarity 46.2%; Pred. No. 88;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 LPAVGLSPGEOE 13
:|:|:|:|:|:|
Db 291 IPAIKGINPDETE 303

RESULT 14

EFGL_PASMU STANDARD; PRT; 700 AA.

AC P57938:

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Elongation factor G (EF-G).

GN FUSA OR PM1356.

OS Pasteurella multocida.

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

OC Pasteurella

OX NCBI_Taxid=747;

RP [1]

RC SEQUENCE FROM N.A.

RA MEDLINE=21145866; PubMed=11248100;

RA May B.J., Zhang O., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;

RT "Complete genomic sequence of Pasteurella multocida pm70.";

RT Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).

CC -1- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT TRANSLLOCATION OF

CC THE NASCENT PROTEIN CHAIN FROM THE A SITE TO THE P-SITE OF THE

CC RIBOSOME.

CC -1- SUBCELLULAR LOCATION: Cytoplasmic.

CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.

CC EF-G/EF-2 SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; AE006173; AK03440.1; -.

DR InterPro: IPR000640; EFG_C.

DR InterPro: IPR000795; GTP_EFTU.

DR InterPro: IPR004161; GTP_EFTU_D2.

DR Pfam: PF00679; EFG_C_1.

DR Pfam: PF00009; GTP_EFTU_1.

DR Pfam: PF03144; GTP_EFTU_D2_1.

DR PRINTS: PR00315; ELONGATNFCF.

DR PROSITE: PS00301; EFATOR_FGP; 1.

KW Elongation factor; Protein biosynthesis; GTP-binding;

KW Complete proteome.

FT NP_BIND 16 23 GTP (BY SIMILARITY).

FT NP_BIND 87 91 GTP (BY SIMILARITY).

FT NP_BIND 141 144 GTP (BY SIMILARITY).

SEQUENCE 700 AA; 77186 MW; 6CC161F7F9FA9C72 CRC64;

Query Match 52.8%; Score 38; DB 1; Length 700;

Best Local Similarity 46.2%; Pred. No. 88;

Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 LPAVGLSPGEOE 13
:|:|:|:|:|:|
Db 292 IPAIKGINPDETE 304

RESULT 15

PMXA_MOUSE

ID PMXA_MOUSE

STANDARD; PRT; 280 AA.

AC O62066;

DT 15-JUL-1999 (Rel. 38, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Paired mesoderm homeobox protein 2A (Paired-like homeobox 2A)

DE (PHOX2A homeodomain protein) (Aristaless homeobox protein homolog).

GN ARX OR PMXA OR PMX2 OR PHOX2A OR PHOX2.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_Taxid=10090;

RP [1]

RC SEQUENCE FROM N.A.

RA MEDLINE=9424481; PubMed=7910552;

RA Valarche T.S., Tissier-Selva J.P., Hirsch M.R., Martinez S.F.,

RA Goridis C., Brunet J.F.;

RT "The mouse homeodomain protein Phox2 regulates Ncam promoter activity

RT in concert with Cux/CDP and is a putative determinant of

RT neurotransmitter phenotype.";

RL Development 119:881-896(1993).

CC -1- FUNCTION: MAY BE INVOLVED IN REGULATING THE SPECIFICITY OF A

CC EXPRESSION OF THE CATECHOLAMINE BIOSYNTHETIC GENES. ACTS AS A

CC TRANSCRIPTION ACTIVATOR/FACTOR. COULD MAINTAIN THE NORADRENERGIC

CC PHENOTYPE (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).

CC -1- SIMILARITY: BELONGS TO THE PAIRED FAMILY OF HOMEBOX PROTEINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; X75014; CA52923.1; -.

DR HESP; P22808; INK3.

DR MGD; MGI:106633; Artx.

DR InterPro: IPR000047; HTH_repressor.

DR InterPro: IPR001356; Homeobox.

DR Pfam; PF00046; homeobox_1.

DR PRINTS: PR00024; HOMEBOX.

DR PRINTS: PR00031; HTHREPRESSR.

DR SMART; SM00389; HOX; 1.

DR PROSITE; PS00027; HOMEBOX_1; 1.

DR PROSITE; PS00071; HOMEBOX_2; 1.

KW Homeobox; DNA-binding; Developmental protein; Nuclear protein.

FT DNA_BIND 90 149 HOMEBOX.

FT DOMAIN 193 198 POLY-PRO.

FT DOMAIN 233 242 POLY-GLY.

SEQUENCE 280 AA; 29417 MW; F7CBFBCE66B57A5D CRC64;

Query Match 51.4%; Score 37; DB 1; Length 280;

Best Local Similarity 58.3%; Pred. No. 52;

Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 LPAVGLSPGEOE 12
:|:|:|:|:|:|
Db 210 LPAALGSGPGPO 221

Search completed: September 9, 2002, 06:56:00
Job time: 333 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2002, 06:49:12 ; Search time 77.59 Seconds
(without alignments)
31.214 Million cell updates/sec

Title: US-09-786-214-12
Perfect score: 72
Sequence: 1 LPAVYGLSPGEQEX 14

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL_19: *
1: sp_archaea: *
2: sp_bacteria: *
3: sp_fungi: *
4: sp_human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_mhc: *
8: sp_organelle: *
9: sp_phage: *
10: sp_plant: *
11: sp_rodent: *
12: sp_virus: *
13: sp Vertebrate: *
14: sp Unclassified: *
15: sp_virus: *
16: sp_bacteriap: *
17: sp_archaeap: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	44	61.1	381	16	O9RRL5	O9rrl5 delnoccuss
2	43	59.7	633	3	O9Y7N6	O9y7n6 schizosacch
3	43	59.7	821	17	O9HPR8	O9hpr8 halobacteri
4	41	56.9	156	16	O9A5K8	O9a5k8 caulobacter
5	41	56.9	228	17	O9HKE3	O9hke3 thermoplasm
6	41	56.9	326	1	O9UXP0	O9uxp0 methanoblu
7	41	56.9	11096	2	O9LAW3	O9law3 streptobus
8	40	55.6	227	17	O9TWD1	O9twd1 sulfobius
9	40	55.6	355	10	O947A7	O947a7 nitellipis
10	40	55.6	358	4	O9NMD0	O9nmd0 homo sapien
11	40	55.6	386	4	O9NMW3	O9nmw3 homo sapien
12	40	55.6	428	16	O9ZAC9	O9zac9 listeria in
13	40	55.6	540	16	O9RR71	O9rr71 delnoccuss
14	40	55.6	661	10	O9SKD1	O9skd1 arabidopsis
15	40	55.6	673	3	O9P327	O9p327 schizosacch
16	40	55.6	683	2	O34302	O34302 thizobium e

17	39	54.2	277	16	O9KOK6	O9kok6 vibrio choi
18	39	54.2	407	2	O9LCW0	O9lcw0 streptomyces
19	38.5	53.5	656	16	O9RW14	O9rw14 delnoccuss
20	38	52.8	136	11	O61061	O61061 mus musculu
21	38	52.8	152	17	O9V252	O9v252 pyrococcus
22	38	52.8	155	17	O57778	O57778 pyrococcus
23	38	52.8	200	16	O97LH7	O97lh7 clostridium
24	38	52.8	212	5	O9NA60	O9na60 caenorhabdi
25	38	52.8	222	17	O9WMC8	O9wmc8 sulfobius
26	38	52.8	280	12	O12652	O12652 broad bean
27	38	52.8	325	16	O9CB52	O9cb52 mycobacteri
28	38	52.8	409	5	O9NSW7	O9nsw7 caenorhabdi
29	38	52.8	519	10	O48655	O48655 oryza sativ
30	38	52.8	540	12	O9IV49	O9iv49 human calic
31	38	52.8	540	12	O918B0	O918b0 norwayk-lik
32	38	52.8	663	9	O94MX4	O94mx4 bacterioph
33	38	52.8	1035	12	O91TP4	O91tp4 lupula herp
34	38	52.8	1317	11	O35211	O35211 mus musculu
35	38	52.8	1769	12	O9WP48	O9wp48 broad bean
36	38	52.8	1804	11	O921Q1	O921q1 mesocricetu
37	38	52.8	1870	12	O9VM10	O9vm10 broad bean
38	38	52.8	2152	2	O9ALM5	O9alm5 saccharopol
39	38	52.8	4823	13	O93321	O93321 fugu rubrip
40	38	52.8	5588	2	O9ALM2	O9alm2 saccharopol
41	38	52.8	6146	2	O93HJ5	O93hj5 streptomyces
42	37.5	52.1	817	2	O53840	O53840 polyanthum
43	37.5	52.1	451	10	O9XHE0	O9xhe0 zea mays (m
44	37.5	52.1	827	11	P70218	P70218 mus musculu
45	37.5	52.1	1804	1	O9P9A9	O9p9a9 uncultured

ALIGNMENTS

RESULT	ID	PRELIMINARY	PRT	381 AA.
O9RRL5	O9RRL5			
AC	O9RRL5			
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	CYTCHROME P450.			
GN	DR2473.			
OS	Delnoccocus radiolurans.			
OC	Bacteria: Thermus/Delnoccocus group; Delnoccocales; Delnoccocus.			
OX	NCBI_TaxID=1299;			
RN	[1]			
RC	SEQUENCE FROM N.A.			
RC	STRAIN-R1;			
RX	MEDLINE-20036896; PubMed-10567266;			
RA	White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,			
RA	Moffat R.J., Haft D.H., Gwin M.L., Nelson W.C., Richardson D.L.,			
RA	Madhavan J.J., Lam P., McDonald L., Uterback T., Zalewski C.,			
RA	Makarova K.A., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,			
RA	Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,			
RA	Fraser C.M.;			
RT	"Genome sequence of the radioreistant bacterium Delnoccocus			
RT	radiolurans R1.";			
RL	Science 286:1571-1577(1999).			
CC	-I- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.			
DR	EMBL; AE002076; AAF12016.1; -.			
DR	TIGR; DR2473; -.			
DR	InterPro: IPR001128; Cyt_P450.			
DR	Pfam: PF00067; P450; 1.			
DR	PRINTS; PRO0385; P450.			
DR	PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.			
KW	Complete proteome; Heme; Monooxygenase; Oxidoreductase.			
SO	SEQUENCE 381 AA; 41940 MW; F191EA69F1797B53 CRC64;			
Query Match	61.1%;	Score 44;	DB 16;	Length 381;
Best Local Similarity	100.0%;	Pred. No. 19;		

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 LPVAVGLSP 9
| | | | | | | | | |
Db 52 LPVAVGLSP 60

RESULT 2
09Y7N6 PRELIMINARY; PRT: 633 AA.

AC 09Y7N6. 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE LYOPHOSPHOLIPASE C1450.09C PRECURSOR (EC 3.1.1.5).
GN SPC1450.09C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
ON NCBI_TaxID=4896;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Lyne M., Rajandream M.A., Barrell B.G., Volckaert G.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CATALYZES THE RELEASE OF FATTY ACIDS FROM
CC LYOPHOSPHOLIPIDS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 2-LYOPHOSPHATIDYLCHOLINE + H(2)O =
CC GLYCEROPHOSPHOCHOLINE + A FATTY ACID ANION.
CC -1- SUBCELLULAR LOCATION: SECRETED (POTENTIAL).
CC -1- SIMILARITY: TO OTHER FUNGAL LYOPHOSPHOLIPASES AND TO YEAST
CC SPO1.
DR EMBL: AL049559; CAB40176.2; -
DR InterPro: IPR002642; PLAC.
DR Pfam: PF01735; PLA2_B; 1.
DR SMART: SM00022; PLAC; 1.
KW Hypothetical protein; Lipid degradation; Hydrolase; Glycoprotein;
KW Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 633 LYOPHOSPHOLIPASE C1450.09C.
FT DOMAIN 404 407 POLY-THR.
FT CARBOHYD 118 118 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 153 153 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 232 232 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 256 256 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 264 264 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 331 331 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 360 360 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 367 367 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 400 400 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 403 403 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 474 474 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 508 508 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 513 513 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 537 537 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 564 564 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 586 586 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 603 603 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 633 AA; 68292 MW; 49871B2955893D19 CRC64;

Query Match 59.7%; Score 43; DB 3; Length 633;
Best Local Similarity 69.2%; Pred. NO. 49;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 2 PAVVGLSPGEQY 14
| | | | | | | | | |
Db 73 PASDGLSTGEQEF 88

RESULT 3
09HPR8 PRELIMINARY; PRT: 821 AA.

AC 09HPR8. 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE DNA HELICASE.
GN HEL OR VNG1501G.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
ON NCBI_TaxID=64091;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.Y., Kennedy S.P., Mahairas G.G., Bergquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Balliga N.S., Thorson V., Sotogona J.,
RA Swartzell S., Weiler D., Hall J., Dahl T.A., Weili R., Goo Y.A.,
RA Leitbauer B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlchroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ebnhardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RT "Genome sequence of Halobacterium species NRC-1.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR EMBL: AE005064; AAG19799.1; -
DR InterPro: IPR001478; PDZ.
DR SMART: SM00228; PDZ; 1.
KW Helicase; Complete proteome.
SQ SEQUENCE 821 AA; 89848 MW; C454C76B98A5702 CRC64;

Query Match 59.7%; Score 43; DB 17; Length 821;
Best Local Similarity 90.0%; Pred. NO. 66;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 3 AVVGLSPGEQ 12
| | | | | | | | | |
Db 326 AVVGLSPAED 335

RESULT 4
09A5K8 PRELIMINARY; PRT: 156 AA.

AC 09A5K8. 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HYPOTHEICAL PROTEIN CC2439.
GN CC2439.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
ON NCBI_TaxID=69394;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19069 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Land M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uitterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.,
RA "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL: AE005913; AAK24410.1; -
DR TIGR: CC2439; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 156 AA; 17142 MW; 21F54D8648396141 CRC64;

Query Match 56.9%; Score 41; DB 16; Length 156;
 Best Local Similarity 80.0%; Pred. No. 24;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 3 AVVGLSPGEQ 12
 DB 18 AVVGLDPGER 27

RESULT 5

ID 09HKE3 PRELIMINARY; PRT; 228 AA.

AC 09HKE3;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
 DE 01-JUN-2001 (TREMBLrel. 17, last annotation update)
 DE PRECORIN-2 METHYLTRANSFERASE RELATED PROTEIN.
 GN TA0658.
 OS Thermoplasma acidophilum.
 OC Archaea; Euryarchaeota; Thermoplasmales; Thermoplasmaceae;
 OC Thermoplasma
 NCBI_TaxID=2303;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DSM 1728;
 RX MEDLINE=20479972; PubMed=11029001;
 RA Ruepp A., Grapl W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
 RA Mews H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
 RT "The genome sequence of the thermophilic scavenger Thermoplasma
 acidophilum."
 RL Nature 407:508-513(2000).
 DR EMBL: AL445065; CAC11796.1;
 DR Interpro: IPR000878; Corrin_porph_mthyltrnf.
 DR Interpro: IPR003043; Uropor_methyltransf.
 DR Pfam: PF00590; TP_methylase; 1.
 DR PROSITE: PS00839; SUMT_1; UNKNOWN_1.
 DR Transferase: Methyltransferase; Complete proteome.
 KW SEQUENCE 228 AA; 25084 MW; 11ABDB86192A67C CRC64;

Query Match 56.9%; Score 41; DB 17; Length 228;
 Best Local Similarity 63.6%; Pred. No. 36;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 4 VVGLSPGEQ 14
 DB 5 VVGLDPGER 15

RESULT 6

ID 09UXP0 PRELIMINARY; PRT; 326 AA.

AC 09UXP0;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
 DE F420-DEPENDENT N5, N10-METHYLENE-TETRAHYDROMETHANOPTERIN REDUCTASE,
 DE PUTATIVE.
 GN FFD.
 OS Methanobacterium tindarius.
 OC Archaea; Euryarchaeota; Methanosarcinales; Methanosarcinaceae;
 OC Methanobacterium
 NCBI_TaxID=2221;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DSM 2278;
 RA Westenberg D.J., Braune A., Ruppert C., Mueller V., Herzberg C.,
 RA Gotschalk G., Blaut M.;
 RT "The F420H₂-dehydrogenase from Methanobacterium tindarius: Cloning of the
 full operon and expression of the genes in Escherichia coli."
 RT Submitted (Sep-1998) to the EMBL/GenBank/DBD databases.
 DR EMBL: M011519; CAB56639.1; -

DR Interpro: IPR002103; Bac_luciferase.
 DR Pfam: PF00296; bac_luciferase; 1.
 SO SEQUENCE 326 AA; 34043 MW; 16F3AB9733A45D82 CRC64;

Query Match 56.9%; Score 41; DB 1; Length 326;
 Best Local Similarity 70.0%; Pred. No. 53;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 3 AVVGLSPGEQ 12
 DB 88 AVVGLDPGER 97

RESULT 7

ID 09LAW3 PRELIMINARY; PRT; 11096 AA.

AC 09LAW3;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
 DE NYSC.
 GN NYSC.
 OS Streptomyces noursei.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
 NCBI_TaxID=1971;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 11455;
 RX MEDLINE=20334850; PubMed=10873841;
 RA Brautaset T., Sekurova O.N., Sletta H., Ellingsen T.E., Strom A.R.,
 RA Valla S., Zotchev S.B.;
 RT "Biosynthesis of the polyene antifungal antibiotic nystatin in
 Streptomyces noursei ATCC 11455: analysis of the gene cluster and
 deduction of the biosynthetic pathway."
 RL Chem. Biol. 7:395-403(2000).
 DR EMBL: AF263912; AAF71776.1; -
 DR HSSP: P25715; IMLA.
 DR Interpro: IPR001227; Acyltransf_domain.
 DR Interpro: IPR002085; Adh_zn_family.
 DR Interpro: IPR000794; Ketoacyl-synt.
 DR Interpro: IPR003880; Phosphopant_attach.
 DR Pfam: PF00698; Acyl_transf; 6.
 DR Pfam: PF00107; adh_zinc; 1.
 DR Pfam: PF00109; ketoacyl-synt; 6.
 DR Pfam: PF02801; ketoacyl-synt_C; 6.
 DR PROSITE: PS0075; ACP_DOMAIN; 6.
 DR PROSITE: PS00606; B_KETOACYL_SYNTHASE; 6.
 DR PROSITE: PS00012; PHOSPHOPANTETHEINE; 5.
 KW Phosphopantetheine; Transferase.
 SO SEQUENCE 11096 AA; 1150415 MW; 776CABAFCAE551DD CRC64;

Query Match 56.9%; Score 41; DB 2; Length 11096;
 Best Local Similarity 63.6%; Pred. No. 2,5e+03;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 2 PAVVGLSPGEQ 12
 DB 4998 PAVVGLDPGER 5008

RESULT 8

ID 097WD1 PRELIMINARY; PRT; 227 AA.

AC 097WD1;
 DT 01-OCT-2001 (TREMBLrel. 18, Created)
 DT 01-OCT-2001 (TREMBLrel. 18, last sequence update)
 DT 01-OCT-2001 (TREMBLrel. 18, last annotation update)
 DE COBALAMIN BIOSYNTHESIS PRECORIN-6B METHYLASE, PUTATIVE
 DE (CBE).
 GN CBE.

OS Sulfolobus solfataricus.
 OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
 OX NCBI_TaxID=2287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 35092 / DSM 1617 / P2;
 RX MEDLINE=21332296; PubMed=11427726;
 RA She O., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
 RA Aweyer M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
 RA De Moers A., Erruso G., Fletcher C., Gordon P.M.K.,
 RA Helkamp-de Jong I., Jeffries A.C., Koza C.J., Medina N., Peng X.,
 RA Thi-Noc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
 RA Takahashi M., Hosoliri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
 RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
 RA Yanamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y.,
 RA Niimiya K., Iwayanagi T.,
 RT "NEDO human cDNA sequencing project."
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK000977; BAA91452.1;
 DR InterPro: IPR003892; CUE.
 DR Pfam: PF02845; CUE. 1.
 DR SMART: SM00546; CUE. 1.
 SO SEQUENCE 227 AA; 25548 MW; 0267F09491F2DCED CRC64;

Query Match 55.6%; Score 40; DB 17; Length 227;
 Best Local Similarity 54.5%; Pred. No. 53;
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 4 VVGLSPGDEY 14
 Db 10 IVGVGPDPDE 20

RESULT 9
 ID 0947A7 PRELIMINARY; PRT; 355 AA.
 AC 0947A7;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE ENOLASE (FRAGMENT).
 GN ENO.
 OS Nitellopsis obtusa.
 OC Eukaryota; Viridiplantae; Streptophyta; Characeae; Nitellopsis.
 OX NCBI_TaxID=40811;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=X-854;
 RX MEDLINE=21437986; PubMed=11526220;
 RA Keeling P.J., Palmer J.D.,
 RT "Lateral transfer at the gene and subgenic levels in the evolution of
 eukaryotic enolase."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:10745-10750(2001).
 DR EMBL; AF348916; AL05455.1;
 FT NON_TER 1
 FT 355 355
 SO SEQUENCE 355 AA; 38295 MW; 400DF160087DE450 CRC64;

Query Match 55.6%; Score 40; DB 10; Length 355;
 Best Local Similarity 54.5%; Pred. No. 87;
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 2 PAVVGLSPGDEQ 12
 Db 28 PAVIGMDPADQ 38

RESULT 10
 ID 09NMD0 PRELIMINARY; PRT; 358 AA.
 AC 09NMD0;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE HYPOTHETICAL 38.8 KDA PROTEIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=EMBRYO;
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Shitatori A., Sudo H.,
 RA Watanabe S., Hosoliri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
 RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
 RA Yanamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y.,
 RA Niimiya K., Iwayanagi T.,
 RT "NEDO human cDNA sequencing project."
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK000977; BAA91452.1;
 DR InterPro: IPR003892; CUE.
 DR Pfam: PF02845; CUE. 1.
 DR SMART: SM00546; CUE. 1.
 SO SEQUENCE 358 AA; 38760 MW; BCE1AA7E95C73BF0 CRC64;

Query Match 55.6%; Score 40; DB 4; Length 358;
 Best Local Similarity 58.3%; Pred. No. 87;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 2 PAVVGLSPGDEQ 13
 Db 213 PAMAGPDPDDE 224

RESULT 11
 ID 09NMW3 PRELIMINARY; PRT; 386 AA.
 AC 09NMW3;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE CDNA FLJ20739 FIS, CLONE HEP07341 (HYPOTHETICAL 42.3 KDA
 PROTEIN).
 GN PROTEIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
 RA Nakajima Y., Mizuno T., Morinaga M., Ota T., Suzuki Y., Obayashi M.,
 RA Nishi T., Shibahara T., Tanaka T., Nakamura Y., Isogai T., Sugano S.,
 RT "NEDO human cDNA sequencing project."
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=MELANOMA;
 RA Strausberg R.,
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK000746; BAA91357.1;
 DR EMBL; BC008034; AAH08034.1;
 DR InterPro: IPR003892; CUE.
 DR Pfam: PF02845; CUE. 1.
 DR SMART: SM00546; CUE. 1.
 SO SEQUENCE 386 AA; 42258 MW; F5530FA47C267895 CRC64;

Query Match 55.6%; Score 40; DB 4; Length 386;
 Best Local Similarity 58.3%; Pred. No. 95;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 2 PAVVGLSPGDEQ 13

Db 213 PMAAGCPEDOE 224

RESULT 12
092AC9 PRELIMINARY; PRT: 428 AA.
AC 092AC9;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
GN PYR PROTEIN.
OS PYR OR LIN1953.
OC Bacteria, Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Listeria.
NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / SEROVAR 6A;
RX PubMed=11679669;

RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Darvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fiehl H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapat G.,
RA Madueno E., Maloumnam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordstlek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Weiland J., Cossart P.,
RT "Comparative genomes of Listeria species."
RL Science 294:849-852(2001).
DR EMBL: AL596170; GAC97183.1; -
DR Listlist: LIN01953; -
KW Complete proteome.
SQ SEQUENCE 428 AA; 44835 MW; 4A2D93B1FA95CCE7 CRC64;

Query Match 55.6%; Score 40; DB 16; Length 428;
Best Local Similarity 70.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 LPAVGLSPG 10
Db 42 VPSVTGLSPG 51

RESULT 13
09RR71 PRELIMINARY; PRT: 540 AA.
AC 09RR71;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
GN CARBOXYLESTERASE, TYPE B.
OS DR2626.
OC Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RL;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Uitterback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus

RT radiodurans RL";
RL Science 286:1571-1577(1999).
CC -1- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.
DR EMBL: AE002092; AAF12163.1; -
DR HSSP: P37967; 1OE3.
DR MEROPS: S09.UNC; -
DR TIGR: DR2626; -
DR InterPro: IPR002018; Carboxylesterase_B.
DR InterPro: IPR000379; Est_Lip_thioest_actite.
DR Pfam: PF00135; Coesterase; 1.
DR PROSITE: PS00122; CARBOXYLESTERASE_B_1; 1.
KW Complete proteome; Hydrolase.
SQ SEQUENCE 540 AA; 55899 MW; CDB97C69E288EFD5 CRC64;

Query Match 55.6%; Score 40; DB 16; Length 540;
Best Local Similarity 70.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 PAVGLSPGE 11
Db 512 PQVGLSPGE 521

RESULT 14
09SKD1 PRELIMINARY; PRT: 661 AA.
AC 09SKD1;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
GN AT2G13900 PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gili J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nielsen W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana."
RL Nature 402:761-768(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC006528; AAF18650.1; -
DR InterPro: IPR004146; DCL.
DR InterPro: IPR001965; PHD.
DR InterPro: IPR001841; Znf_ring.
DR Pfam: PF03107; DCL; 2.
DR SMART: SM00249; PHD; 4.
DR SMART: SM00184; RING; 2.
SQ SEQUENCE 661 AA; 76822 MW; 5CE1FD9F848A60D7 CRC64;

Query Match 55.6%; Score 40; DB 10; Length 661;
Best Local Similarity 53.8%; Pred. No. 1.7e+02;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 2 PAVGLSPGE 14

Db 437 PLTKIVGEGEX 449

RESULT 15

09P327 PRELIMINARY; PRT; 673 AA.

AC 09P327;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE PUTATIVE LYOPHOSPHOLIPASE PRECURSOR.
 GN SPAC1348.10C OR SPAC977.09C
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 CC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972H-;
 RA Lyne M., Barrell B.G., Rajandream M.A.;
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972H-;
 RA Zimmermann W., Wambutt R., McDougall R.C., Rajandream M.A.,
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL358912; CAB94277.1; -.
 DR EMBL; AL137130; CAB69631.1; -.
 DR InterPro; IPR001179; FKBP_PPIase.
 DR InterPro; IPR002642; PLAC.
 DR Pfam; PF01735; PLA2_B; 1.
 DR SMART; SM00453; FKBP_PPIase_1; UNKNOWN_1.
 DR PROSITE; PS00453; FKBP_PPIase_1; UNKNOWN_1.
 SO SEQUENCE 673 AA: 74595 MW: B39A773E76CD694B CRC64;

Query Match

Best Local Similarity 55.6%; Score 40; DB 3; Length 673;
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 2 PAVVGLSPGEGEX 14
 11 11:1111
 Db 83 PASEGLNEGEGSY 95

Search completed: September 9, 2002, 06:55:26
 Job time: 374 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2002, 06:36:12 : Search time 64.82 seconds
(without alignments)
23.990 Million cell updates/sec

Title: US-09-786-214-12
Perfect score: 72
Sequence: 1 LPAYGLSPGEQY 14

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 747574 seqs, 111073796 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: A.Geneseq_032802.*
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	72	100.0	14	21	AAV84266
2	72	100.0	20	21	AAV84265
3	72	100.0	25	21	AAV84264
4	68	94.4	13	21	AAV84267
5	65	90.3	13	21	AAV84268
6	65	90.3	15	21	AAV84269
7	45	62.5	234	22	AAV84268
8	41	56.9	306	22	AAV84268
9	41	56.9	11096	22	AAV84268
10	40	55.6	361	22	AAV84268
11	40	55.6	306	22	AAV84268

12	40	55.6	358	22	AAV84266
13	40	55.6	668	21	AAV84265
14	39	54.2	18	20	AAV84264
15	39	54.2	18	22	AAV84267
16	39	54.2	18	22	AAV84268
17	39	54.2	18	22	AAV84269
18	39	54.2	88	21	AAV84268
19	39	54.2	104	18	AAV84268
20	39	54.2	104	18	AAV84268
21	39	54.2	106	14	AAV84268
22	39	54.2	106	14	AAV84268
23	39	54.2	107	14	AAV84268
24	39	54.2	107	15	AAV84268
25	39	54.2	107	15	AAV84268
26	39	54.2	107	18	AAV84268
27	39	54.2	107	18	AAV84268
28	39	54.2	107	18	AAV84268
29	39	54.2	107	18	AAV84268
30	39	54.2	107	18	AAV84268
31	39	54.2	107	18	AAV84268
32	39	54.2	107	19	AAV84268
33	39	54.2	107	19	AAV84268
34	39	54.2	107	20	AAV84268
35	39	54.2	107	20	AAV84268
36	39	54.2	107	20	AAV84268
37	39	54.2	107	20	AAV84268
38	39	54.2	107	21	AAV84268
39	39	54.2	107	21	AAV84268
40	39	54.2	107	22	AAV84268
41	39	54.2	107	22	AAV84268
42	39	54.2	107	22	AAV84268
43	39	54.2	107	22	AAV84268
44	39	54.2	107	22	AAV84268
45	39	54.2	107	22	AAV84268

ALIGNMENTS

RESULT: 1
ID: AAV84266 standard; Peptide: 14 AA.
AC: AAV84266; (first entry)
XX: 12-JUL-2000
DE: Peptide derived from macrophage colony stimulating gene alternative ORF.
XX: tumour rejection antigen; macrophage colony stimulating gene;
KW: macrophage-colony stimulating factor; antigen presenting cell;
KW: human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
OS: Synthetic.
OS: Homo sapiens.
PN: WO200013699-A1
XX: 16-MAR-2000.
PD: 03-SEP-1999; 99WO-US20344.
XX: 04-SEP-1998; 98US-0099077.
XX: (LUDW-) LUDWIG INST CANCER RES.
XX: Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;
XX: WPI; 2000-256859/22.
XX: Isolated polypeptide used to treat subjects having a disorder
XX: characterized by expression of alternative open reading frame
XX: macrophage-colony stimulating factor comprises 25 amino acid residue

PT sequence -
XX
PS Claim 2; Page 39; 74pp; English.
XX
CC The present sequence represents a peptide which is derived from a tumour
CC rejection antigen precursor encoded by an alternative open reading frame
CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC the alternative ORF of macrophage-colony stimulating factor, when
CC presented by an antigen presenting cell having a human leukocyte antigen
CC (HLA) class I molecule, effectively induce the activation and
CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC acids derived from the alternate ORF of macrophage-colony stimulating
CC factor are useful for enriching selectively a population of T lymphocytes
CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC characterized by expression of the polypeptide, and for identifying
CC functional variants and mimetics.
XX
SQ Sequence 14 AA;
XX
Query Match 100.0%; Score 72; DB 21; Length 14;
Best Local Similarity 100.0%; Pred. No. 3e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LPAVGLSPGEQEX 14
1 |
Db 1 lpavglspgeqey 14
XX
RESULT 2
AAY84265
ID AAY84265 standard; Peptide: 20 AA.
XX
AC AAY84265;
XX
DT 12-JUL-2000 (first entry)
XX
DE Truncated macrophage colony stimulating factor tumour antigen.
XX
KW tumour rejection antigen; macrophage colony stimulating gene;
KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS Homo sapiens.
XX
PN WO200013699-A1.
XX
PD 16-MAR-2000.
XX
PE 03-SEP-1999; 99WO-US20344.
XX
PR 04-SEP-1998; 98US-0099077.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Probst-Keppler M, Van Den Eynde B, Boon-Falleur T;
XX
PS WPI: 2000-256859/22.
XX
DR N-PSDB: AA299675.
XX
PT Isolated polypeptide used to treat subjects having a disorder
PT characterized by expression of alternative open reading frame
PT macrophage-colony stimulating factor comprises 25 amino acid residue
PT sequence -
XX
PS Claim 3; Page 64; 74pp; English.
XX
CC The present sequence represents a truncated tumour rejection antigen
CC precursor, and is encoded by a truncated alternative open reading frame
CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC the alternative ORF of macrophage-colony stimulating factor, when
CC presented by an antigen presenting cell having a human leukocyte antigen
CC (HLA) class I molecule, effectively induce the activation and

CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC acids derived from the alternate ORF of macrophage-colony stimulating
CC factor are useful for enriching selectively a population of T lymphocytes
CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC characterized by expression of the polypeptide, and for identifying
CC functional variants and mimetics.
XX
SQ Sequence 20 AA;
XX
Query Match 100.0%; Score 72; DB 21; Length 20;
Best Local Similarity 100.0%; Pred. No. 4.5e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LPAVGLSPGEQEX 14
1 |
Db 4 lpavglspgeqey 17
XX
RESULT 3
AAY84264
ID AAY84264 standard; Peptide: 25 AA.
XX
AC AAY84264;
XX
DT 12-JUL-2000 (first entry)
XX
DE Peptide of alternate reading frame of macrophage colony stimulating gene.
XX
KW Renal cell carcinoma; antigen; cytotoxic T lymphocyte;
KW tumour rejection antigen; macrophage colony stimulating gene;
KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS Homo sapiens.
XX
PN WO200013699-A1.
XX
PD 16-MAR-2000.
XX
PE 03-SEP-1999; 99WO-US20344.
XX
PR 04-SEP-1998; 98US-0099077.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Probst-Keppler M, Van Den Eynde B, Boon-Falleur T;
XX
PS WPI: 2000-256859/22.
XX
DR N-PSDB: AA299672.
XX
PT Isolated polypeptide used to treat subjects having a disorder
PT characterized by expression of alternative open reading frame
PT macrophage-colony stimulating factor comprises 25 amino acid residue
PT sequence -
XX
PS Claim 1; Page 64; 74pp; English.
XX
CC The present sequence represents a tumour rejection antigen precursor,
CC and is encoded by an alternative open reading frame (ORF) of human
CC macrophage colony stimulating gene. Peptides derived from the
CC alternative ORF of macrophage-colony stimulating factor, when presented
CC by an antigen presenting cell having a human leukocyte antigen (HLA)
CC class I molecule, effectively induce the activation and proliferation
CC of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic acids
CC derived from the alternate ORF of macrophage-colony stimulating factor
CC are useful for enriching selectively a population of T lymphocytes
CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC characterized by expression of the polypeptide, and for identifying
CC functional variants and mimetics.
XX
SQ Sequence 25 AA;

Query Match 100.0%; Score 72; DB 21; Length 25;
 Best Local Similarity 100.0%; Pred. No. 5,7e-05;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LPAVGLSPGEQEX 14
 |||||
 DB 4 LPavglspgeqey 17

RESULT 4

ID AAY84267 standard; Peptide: 13 AA.

AC AAY84267;

DT 12-JUL-2000 (first entry)

DE Peptide derived from macrophage colony stimulating gene alternative ORF.

KW tumour rejection antigen; macrophage colony stimulating gene;

KM macrophage-colony stimulating factor; antigen presenting cell;

XX human leukocyte antigen; CD8+ cytotoxic T lymphocyte.

OS Synthetic.

OS Homo sapiens.

PN WO200013699-A1

PD 16-MAR-2000

PF 03-SEP-1999; 99WO-US20344.

PR 04-SEP-1998; 98US-0099077.

PA (LUDW-) LUDWIG INST CANCER RES.

PI Probst-Keppler M, Van Den Eynde B, Boon-Falleur T;

DR WPI; 2000-256859/22.

PT Isolated polypeptide used to treat subjects having a disorder
 PT characterized by expression of alternative open reading frame
 PT macrophage-colony stimulating factor comprises 25 amino acid residue
 sequence

PS Example 2; Page 40; 74pp; English.

CC The present sequence represents a peptide which is derived from a tumour
 CC rejection antigen precursor encoded by an alternative open reading frame
 CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
 CC the alternative ORF of macrophage-colony stimulating factor, when
 CC presented by an antigen presenting cell having a human leukocyte antigen
 CC (HLA) class I molecule, effectively induce the activation and
 CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
 CC acids derived from the alternate ORF of macrophage-colony stimulating
 CC factor are useful for enriching selectively a population of T lymphocytes
 CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
 CC characterized by expression of the polypeptide, and for identifying
 CC functional variants and mimetics.

CC XX Sequence 13 AA;

Query Match 94.4%; Score 68; DB 21; Length 13;

Best Local Similarity 100.0%; Pred. No. 0.00012;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 PAVGLSPGEQEX 14

DB 1 pavglspgeqey 13

RESULT 5
 ID AAY84268 standard; Peptide: 13 AA.

AC AAY84268;

DT 12-JUL-2000 (first entry)

DE Peptide derived from macrophage colony stimulating gene alternative ORF.

KW tumour rejection antigen; macrophage colony stimulating gene;

KM macrophage-colony stimulating factor; antigen presenting cell;

XX human leukocyte antigen; CD8+ cytotoxic T lymphocyte.

OS Synthetic.

OS Homo sapiens.

PN WO200013699-A1

PD 16-MAR-2000

PF 03-SEP-1999; 99WO-US20344.

PR 04-SEP-1998; 98US-0099077.

PA (LUDW-) LUDWIG INST CANCER RES.

PI Probst-Keppler M, Van Den Eynde B, Boon-Falleur T;

DR WPI; 2000-256859/22.

PT Isolated polypeptide used to treat subjects having a disorder
 PT characterized by expression of alternative open reading frame
 PT macrophage-colony stimulating factor comprises 25 amino acid residue
 sequence

PS Example 2; Page 40; 74pp; English.

CC The present sequence represents a peptide which is derived from a tumour
 CC rejection antigen precursor encoded by an alternative open reading frame
 CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
 CC the alternative ORF of macrophage-colony stimulating factor, when
 CC presented by an antigen presenting cell having a human leukocyte antigen
 CC (HLA) class I molecule, effectively induce the activation and
 CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
 CC acids derived from the alternate ORF of macrophage-colony stimulating
 CC factor are useful for enriching selectively a population of T lymphocytes
 CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
 CC characterized by expression of the polypeptide, and for identifying
 CC functional variants and mimetics.

CC XX Sequence 13 AA;

Query Match 90.3%; Score 65; DB 21; Length 13;

Best Local Similarity 100.0%; Pred. No. 0.00039;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LPAVGLSPGEQEX 13

DB 1 LPavglspgeqe 13

RESULT 6

ID AAY84269 standard; Peptide: 15 AA.

AC AAY84269;

DT 12-JUL-2000 (first entry)

DE Peptide derived from macrophage colony stimulating gene alternative ORF.

KW tumour rejection antigen; macrophage colony stimulating gene;
KM macrophage-colony stimulating factor; antigen presenting cell;
XX human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
OS Synthetic.
XX Homo sapiens.
XX WO200013699-A1.
PD 16-MAR-2000.
XX PF 03-SEP-1999; 99WO-US20344.
XX PR 04-SEP-1998; 98US-0099077.
XX PA (LUDW-) LUDWIG INST CANCER RES.
XX PI Probst-Keeper M, Van Den Eynde B, Boon-Falleur T;
XX DR WPI; 2000-256859/22.
XX PT Isolated polypeptide used to treat subjects having a disorder
XX PT characterized by expression of alternative open reading frame
XX PT macrophage-colony stimulating factor comprises 25 amino acid residue
XX PT sequence -
XX PS Example 2: Page 40; 74pp; English.
XX CC The present sequence represents a peptide which is derived from a tumour
XX CC rejection antigen precursor encoded by an alternative open reading frame
XX CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
XX CC the alternative ORF of macrophage-colony stimulating factor, when
XX CC presented by an antigen presenting cell having a human leukocyte antigen
XX CC (HLA) class I molecule, effectively induce the activation and
XX CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
XX CC acids derived from the alternate ORF of macrophage-colony stimulating
XX CC factor are useful for enriching selectively a population of T lymphocytes
XX CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
XX CC characterized by expression of the polypeptide, and for identifying
XX CC functional variants and mimetics.
XX SQ Sequence 15 AA;
SQ
Query Match 90.3%; Score 65; DB 21; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00045;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LPVAVGLSPGEQ 13
Db 3 lpavvglspege 15
RESULT 7
AAB36208
ID AAB36208 standard; protein; 234 AA.
XX
XX AAB36208;
XX
XX 15-FEB-2001 (first entry)
XX
XX Human immune system associated protein HISAP-6.
XX DE Human; immune system associated protein; HISAP-6; immune disorder;
XX KW infection; autoimmune disease; cancer.
XX
XX Homo sapiens.
XX OS
XX US6135941-A.
XX PN 24-OCT-2000.
XX PD 27-MAR-1998; 98US-0049672.
XX PF

XX
XX 27-MAR-1998; 98US-0049672.
XX
XX (INCY-) INCYTE PHARM INC.
XX PA
XX PI Tang YT, Yue H, Lal P, Corley NC, Guegler KJ, Baughn MR;
XX PI Hallman JL, Au-Young J;
XX
XX WPI: 2001-030926/04.
XX DR N-PSDB; AAC66524.
XX
XX New human immune system associated proteins (HISAP) and polynucleotides
XX PT encoding the HISAP, useful for diagnosing, treating or preventing
XX PT immune or cell proliferative disorders or infections -
XX PS Claim 1; Column 59-60; 54pp; English.
XX
XX The present invention provides the coding and protein sequences for a
XX CC number of human immune system associated proteins (HISAPs). These can be
XX CC used in the diagnosis and treatment of various autoimmune disorders,
XX CC infections and cell proliferative diseases. The diseases include AIDS,
XX CC adult respiratory distress syndrome, anaemia, asthma, atherosclerosis,
XX CC Crohn's disease, irritable bowel syndrome, multiple sclerosis, myasthenia
XX CC gravis, osteoarthritis, rheumatoid arthritis, scleroderma, systemic lupus
XX CC erythematosus, arteriosclerosis, cirrhosis and cancer.
XX
XX Sequence 234 AA;
SQ
SQ
Query Match 62.5%; Score 45; DB 22; Length 234;
Best Local Similarity 72.7%; Pred. No. 17;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 2 PAVVGLSPGEQ 12
Db 28 pavislspegr 38
RESULT 8
ID ABB19552
ABG19552 standard; protein; 306 AA.
XX
XX ABB19552;
XX AC
XX 13-FEB-2002 (first entry)
XX DT
XX DE Novel human diagnostic protein #19543.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX KW
XX Homo sapiens.
XX OS
XX WO200175067-A2.
XX PN
XX 11-OCT-2001.
XX PD
XX 30-MAR-2001; 2001WO-US08631.
XX PF
XX 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX PA
XX Drmanac RT, Liu C, Tang YT;
XX PI
XX WPI: 2001-639362/73.
XX DR N-PSDB; AAS83739.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity -
XX PT

XX Claim 20: SEQ ID No 49911; 103bp; English.
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations in
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WtPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 SQ Sequence 306 AA:

Query Match 56.9%; Score 41; DB 22; Length 306;
 Best Local Similarity 57.1%; Pred. No. 1e+02;
 Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Oy 1 LPVWGLSPGQEV 14
 Db 137 LPQAPGISPGQSW 150

RESULT 9
 ID AAE10129 standard; Protein; 11096 AA.
 AC AAE10129;
 DT 29-NOV-2001 (first entry)
 DE Streptomyces noursei nystatin gene, MYSC.
 KW Polypeptide synthase; PKS; macrolide; nystatin; PKS gene cluster;
 XX antifungal; antibiologic; PKS type I.
 OS Streptomyces noursei.
 FH Key
 FT Location/Qualifiers
 FT 35..455
 FT /label= KS3_domain
 FT /note= "ketosynthase (KS) domain"
 FT 546..858
 FT /label= AR3_domain
 FT /note= "acyltransferase (AT) domain"
 FT 872..1073
 FT /label= DH3_domain
 FT /note= "dehydratase (DH) domain"
 FT 1381..1628
 FT /label= KR3_domain
 FT /note= "ketoreductase (KR) domain"
 FT 1662..1735
 FT /label= ACP3_domain
 FT /note= "Acyl carrier protein (ACP) domain"
 FT 1757..2180
 FT /label= KS4_domain
 FT /note= "ketosynthase (KS) domain"
 FT 2291..2603
 FT /label= AT4_domain

FT /note= "acyltransferase (AT) domain"
 FT 2617..2818
 FT /label= DH4_domain
 FT /note= "dehydratase (DH) domain"
 FT 3124..3371
 FT /label= KR4_domain
 FT /note= "ketoreductase (KR) domain"
 FT 3407..3480
 FT /label= ACP4_domain
 FT /note= "Acyl carrier protein (ACP) domain"
 FT 3501..3924
 FT /label= KS5_domain
 FT /note= "ketosynthase (KS) domain"
 FT 4032..4346
 FT /label= AT5_domain
 FT /note= "acyltransferase (AT) domain"
 FT 4360..4561
 FT /label= DH5_domain
 FT /note= "dehydratase (DH) domain"
 FT 4953..5239
 FT /label= ER5_domain
 FT /note= "enoylreductase (ER) domain"
 FT 5248..5495
 FT /label= KR5_domain
 FT /note= "ketoreductase (KR) domain"
 FT 5528..5601
 FT /label= ACP5_domain
 FT /note= "Acyl carrier protein (ACP) domain"
 FT 5623..6046
 FT /label= KS6_domain
 FT /note= "ketosynthase (KS) domain"
 FT 6166..6478
 FT /label= AT6_domain
 FT /note= "acyltransferase (AT) domain"
 FT 6492..6704
 FT /label= DH6_domain
 FT /note= "dehydratase (DH) domain"
 FT 7038..7281
 FT /label= KR6_domain
 FT /note= "ketoreductase (KR) domain"
 FT 7315..7388
 FT /label= ACP6_domain
 FT /note= "Acyl carrier protein (ACP) domain"
 FT 7408..7831
 FT /label= KS7_domain
 FT /note= "ketosynthase (KS) domain"
 FT 7939..8253
 FT /label= AT7_domain
 FT /note= "acyltransferase (AT) domain"
 FT 8267..8470
 FT /label= DH7_domain
 FT /note= "dehydratase (DH) domain"
 FT 8812..9086
 FT /label= KR7_domain
 FT /note= "ketoreductase (KR) domain"
 FT 9120..9193
 FT /label= ACP7_domain
 FT /note= "Acyl carrier protein (ACP) domain"
 FT 9214..9637
 FT /label= KS8_domain
 FT /note= "ketosynthase (KS) domain"
 FT 9758..10072
 FT /label= AT8_domain
 FT /note= "acyltransferase (AT) domain"
 FT 10086..10289
 FT /label= DH8_domain
 FT /note= "dehydratase (DH) domain"
 FT 10657..10904
 FT /label= KR8_domain
 FT /note= "ketoreductase (KR) domain"
 FT 10939..11012
 FT /label= ACP8_domain
 FT /note= "Acyl carrier protein (ACP) domain"

```

XX      16-AUG-2001.
PN      WO200159126-A2.
XX
XX      08-FEB-2001; 2001WO-GB00509.
PD
XX
XX      08-FEB-2000; 2000GB-0002840.
XX      10-APR-2000; 2000GB-0008786.
XX      14-APR-2000; 2000GB-0009387.
PR
XX
XX      (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELLIGE.
PA      (SMTF) SINTEF STIFTELSEN IND TEK FORSK.
PA      (ALPH-) ALPHARMA AS.
PA      (SINV-) SINVENT AS.
PA      (DZIE/) DZIELEWSKA H.
PA      (ZOTC/) ZOTCHEV S B.
PA      (SEKU/) SEKUROVA O N.
PA      (FJAE/) FJAEVYIK E.
PA      (BRAU/) BRAUTASET T.
PA      (STRO/) STROM A R.
XX
XX      ZOTCHEV SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
PI      Valla S, Ellingsen TE, Sielta H, Gulliksen O;
XX
XX      WPI; 2001-557614/62.
DR      N-PSDB; AAD17184.
XX
XX      New nystatin polyketide synthase polynucleotides and polypeptides,
PT      useful as antibiotics and antifungals -
XX
XX      Claim 15; Page 170-176; 266pp; English.
XX
XX      The present invention relates to the cloning and sequencing of the gene
CC      cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC      involved in the biosynthesis of the macrolide antibiotic nystatin.
CC      The nystatin PKS is useful as antifungal antibiotics. The present
CC      sequence is a PKS type I encoding Streptomyces noursei nystatin
CC      gene, NysC.
XX
XX      Sequence 11096 AA;
SQ

```

Query Match 56.9%; Score 41; DB 22; Length 11096;
 Best Local Similarity 63.6%; Pred. No. 5.4e+03;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```

OY      2 PAVVGLSPGEQ 12
DB      4998 pevgtglapgdq 5008

```

RESULT 10
 AAG91441
 ID AAG91441 standard; Protein: 261 AA.
 XX
 AC AAG91441:
 XX
 DT 26-SEP-2001 (first entry)
 XX
 DE C glutamicum protein fragment SEQ ID NO: 5195.
 XX
 KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
 KW organic acid synthesis.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN EP1108790-A2.
 XX
 PD 20-JUN-2001.
 XX
 PF 18-DEC-2000; 2000EP-0127688.
 XX

```

PR      16-DEC-1999; 99JP-0377484.
PR      07-APR-2000; 2000JP-0159162.
PR      03-AUG-2000; 2000JP-0280988.
XX
XX      (KYOW ) KYOWA HAKKO KOGYO KK.
XX
XX      Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI      Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
XX      WPI; 2001-376931/40.
DR      N-PSDB; AAH66660.
XX
XX      Novel polynucleotides derived from Coryneform bacteria, for identifying
PT      mutation point of a gene, measuring expression of a gene, analysing
PT      expression profile or pattern of a gene and identifying homologous gene
XX
XX      Claim 17; SEQ ID NO: 5195; 246pp + Sequence Listing; English.
XX
XX      The present invention provides a number of nucleotide and protein
CC      sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC      are useful for identifying the mutation point of a gene derived from a
CC      mutant of coryneform bacterium, measuring expression amount and
CC      analysing the expression profile or expression pattern of a gene derived
CC      from Coryneform bacterium, and identifying a homologue of a gene derived
CC      from coryneform bacterium. Coryneform bacteria are useful for producing
CC      amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC      particularly L-lysine. The present sequence is a protein described
CC      in the exemplification of the invention.
CC      Note: The sequence data for this patent did not form part of the printed
CC      specification, but was obtained in electronic format directly from the
CC      European Patent Office.
XX
XX      Sequence 261 AA;
SQ

```

Query Match 55.6%; Score 40; DB 22; Length 261;
 Best Local Similarity 53.8%; Pred. NO. 1.3e+02;
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

```

OY      1 LPAVGLSPGEQ 13.
DB      137 lpalatvpsgead 149

```

RESULT 11
 ABG24698
 ID ABG24698 standard; Protein: 306 AA.
 XX
 AC ABG24698:
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #24689.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEO INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX

PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
XX
PI Rosen CA, Ruben SM;
XX
XX WPI: 2000-587513/55.
DR N-PSDB; AAF16258.
XX
PT Prostate cancer associated gene sequences, referred to as prostate
PT cancer antigens, useful for treatment, prevention, and diagnosis of
PT disorders such as prostate cancer -
XX
XX
PS Claim 11: Page 2091-2094; 2338pp; English.
XX
XX AAF15566 to AAF16505 encode the human prostate cancer associated
CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC The prostate cancer antigens can have neuroprotective, cytostatic,
CC cardiolactive, immunomodulatory, muscular, vulnary, gastrointestinal,
CC nephrotropic, antineoplastic, gynaecological and antibacterial activities,
CC and can be used in gene therapy. The prostate cancer antigen
CC polynucleotides may be used for detection of prostate cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The prostate cancer antigens may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
CC AAF57303 represent sequences used in the exemplification of the present
CC invention.
XX
XX
SQ Sequence 668 AA;

Query Match 55.6%; Score 40; DB 21; Length 668;
Best Local Similarity 72.7%; Pred. No. 3.6e+02;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 4 VVGLSPGEQY 14
|||||||:
Db 27 vvglspsgyfy 37

RESULT 14
AAV41875
ID AAV41875 standard; Peptide: 18 AA.
XX
XX AAV41875;
XX
XX 09-DEC-1999 (first entry)
XX
XX Rheumatoid arthritis diagnostic protein isoform peptide #26.
XX
XX Human: rheumatoid arthritis; RA; diagnosis; RPI; RADF; detection;
KW Rheumatoid arthritis diagnostic feature; RPI; synovial fluid;
KW Rheumatoid arthritis diagnostic protein isoform; screening;
KW expression reference protein isoform; prognosis.
XX
XX Homo sapiens.
XX
XX OS
XX PN W09947925-A2.
XX PD 23-SEP-1999.
XX
XX PF 15-MAR-1999; 99WO-GB00763.
XX
XX PR 13-MAR-1998; 98GB-0005477.
XX
XX (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
XX PI Parekh RB, Patel TP, Townsend RR;
XX
XX WPI: 1999-571871/48.
XX
XX
PT Diagnosis of human rheumatoid arthritis by two-dimensional

PT electrophoresis -
XX
XX
PS Disclosure; Page 18; 157pp; English.
XX
XX

CC A method has been developed for the diagnosis of human rheumatoid
CC arthritis (RA) using two-dimensional electrophoresis to generate a
CC two-dimensional array of features. The method can be used for screening,
CC diagnosis and prognosis of RA in a subject or for monitoring the effect
CC of an anti-RA drug or therapy administered to a subject. The method
CC comprises: (a) analysing a sample of serum or plasma and optionally
CC synovial fluid by two-dimensional electrophoresis, to generate a two-
CC dimensional array of features; (b) identifying at least one chosen
CC feature whose relative abundance correlates with the presence or absence
CC of RA; and (c) comparing the abundance of each chosen feature in the
CC sample with the abundance of that chosen feature in serum or plasma from
CC one or more persons without RA, where the relative abundance of the
CC chosen feature or features in the sample indicates the presence or
CC absence of RA in the subject. The method can also be used in clinical
CC studies for testing drugs for therapy of RA, for purification of RA-
CC diagnostic protein isoforms (RPIs), and for production of antibodies to
CC RPIs. The RA-diagnostic feature (RADF) proteins can be used to identify
CC compounds that promote or inhibit their activity, which are then used as
CC RA drugs. Nucleic acid encoding RADFs can be used in gene therapy
CC protocols. AAV41844 to AAV42100 represent RPI peptides, AAV42101 to
CC AAV42103 represent expression reference protein isoform peptides and
CC AAV25066 to AAV25068 represent degenerate probes for RPIs, which are all
CC used in the exemplification of the present invention.
XX
XX
SQ Sequence 18 AA;

Query Match 54.2%; Score 39; DB 20; Length 18;
Best Local Similarity 63.6%; Pred. No. 9.7;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 2 PAVGLSPGEQ 12
||:|||||:
Db 8 patlspsger 18

RESULT 15
AAU25388
ID AAU25388 standard; Peptide: 18 AA.
XX
XX AAU25388;
XX
XX 18-DEC-2001 (first entry)
XX
XX Schizophrenia-Associated Protein Isoform (SPI) peptide #617.
XX
XX Schizophrenia-associated protein isoform; SPI; SPI-206; SPI-238; SPI-240;
KW neuroleptic; gene therapy; cerebrospinal fluid; serum; plasma.
XX
XX Homo sapiens.
XX
XX OS
XX PN W0200162785-A2.
XX PD 30-AUG-2001.
XX
XX PF 23-FEB-2001; 2001WO-GB00792.
XX
XX PR 24-FEB-2000; 2000GB-0004415.
XX
XX PR 28-NOV-2000; 2000US-0750395.
XX
XX (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
XX PI Herath HMA, Parekh RB, Rohlf C, Terrett JA, Tyson KL;
XX
XX WPI: 2001-570624/64.
XX
XX
XX New schizophrenia associated protein isoforms and encoding nucleic acid
PT molecules, useful for treatment, diagnosis and prognosis of
PT schizophrenia and screening for potential drugs for treatment and new

PT drug targets -

XX
PS Disclosure: Page 41; 148pp; English.
XXCC The sequence represents a schizophrenia-associated protein isoform (SPI).
CC These protein isoforms, e.g. SPI-206, SPI-238 and SPI-240 are detectable
CC in cerebrospinal fluid, serum or plasma and are useful markers of
CC schizophrenia. The sequences can be used for treatment and diagnosis of
CC schizophrenia, screening, prognosis, monitoring the results of therapy,
CC identifying patients most likely to respond to a particular therapy and
CC identification of new targets for drug treatment. SPI DNA is useful as a
CC nucleic acid probe to detect the presence of nucleic acids or SPIs.
XX

SQ Sequence 18 AA:

Query Match 54.2%; Score 39; DB 22; Length 18;
Best Local Similarity 63.6%; Pred. NO. 9.7;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;OY 2 PAVVGLSPGEQ 12
II : II II I I :
Db 8 patlslspger 18Search completed: September 9, 2002, 06:50:21
Job time: 849 sec

